Mathematical Modeling and Experimental Verification of Resonance Energy Transfer Networks: Applications in Cryptography and Biological Sensing

by

Vishwa Nellore

Department of Electrical and Computer Engineering
Duke University

Date: 11/24/2014
Approved:

_____
Christopher Dwyer, Supervisor

_____
Alvin Lebeck

_____
Jeff Glass

_____
John Reif

_____
Ravikanth Pappu

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Electrical and Computer Engineering in the Graduate School
of Duke University

2014

ABSTRACT

Mathematical Modeling and Experimental Verification of Resonance Energy Transfer Networks: Applications in Cryptography and Biological Sensing

by

Vishwa Nellore

Department of Electrical and Computer Engineering
Duke University

Date: 11/24/2014
Approved:

_____
Christopher Dwyer, Supervisor

_____
Alvin Lebeck

_____
Jeff Glass

_____
John Reif

_____
Ravikanth Pappu

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Electrical and Computer Engineering in the Graduate School of
Duke University

2014

Abstract

FRET between pairs of fluorophores is widely used as a biological assay. However, the properties of larger fluorophore networks are poorly understood and their application space has not yet been fully explored. This dissertation introduces DNA self-assembled Förster Resonance Energy Transfer (FRET) networks that provide a unique optical output when probed by a series of light pulses. We create a Markov model of the FRET networks and analyze over 1200 time-resolved fluorescence measurements on 108 unique prototypical networks. Our results show that the optical responses of FRET networks are highly repeatable and minor variations between the FRET networks can be discriminated resulting in a total of 10^{375} unique responses. These results are used in the following breakthrough applications:

1. Unclonable Cryptographic Key for Secure Authentication

Modern authentication protocols rely on an asymmetry in the effort required by a legitimate user and an adversary to accurately decrypt an encoded message. These protocols ensure that communication between legitimate users is possible in polynomial time using a private key but a user without access to the exact key cannot compute the function using a probabilistic polynomial-time algorithm. Private-key cryptographic techniques currently employ physical keys based on algorithmic one-way functions, which are conjectured mathematical objects that are easy to compute but difficult to

invert. Well-known examples of such one-way functions include the RSA and the Rabin functions. Although algorithmic one-way functions are widely used for authentication, their reliance on computational difficulty to provide security implies that they are not protected against future advances in computational capacity or speed. Also, use of a highly parallel network of conventional computers could potentially reverse engineer a key from the challenge-response pairs used in past communications. The key may also be obtained by duplicating the device. Most of the current physical embodiments of algorithmic one-way functions come with a tamper resistant packaging but remain vulnerable to sophisticated attacks.

We develop a RET based physical key to overcome the limitations of conventional security keys. The key exploits resonance energy transfer between a network of fluorophores placed on a nanostructure. The fluorophores provide a unique, unpredictable output when probed by a series of light pulses of specific wavelengths and delays. A critical advantage of the RET key over existing keys is that the manufacturing process allows two identical devices to be produced allowing us to exploit the advantages of symmetric key encryption, for the first time, without the need to physically transfer the device between the two communicating parties.
It is infeasible to model, characterize or replicate our key using modern cryptographic attacks including unfettered physical access to the device. This is because of the difficulty in characterizing the nanoscale structure and the large number of challenge-

response pairs achievable for each key. Atomic force microscopy and time –resolved fluorescence measurements are performed to characterize the nanoscale structures. From over 1200 measurements on 108 unique prototypical keys, we estimate that a legitimate user would have a computational advantage of 10^{340} years over an attacker even if the attacker uses all the computational resources available in the world. Thus, the computational advantage of our key ensures perfect theoretical security for the foreseeable future. We provide an authentication protocol for use of the key and demonstrate that legitimate users are successfully authenticated 99.48% of the time with two trials.

2. Multiplexed Fluorescence Sensor for Cancer Detection

Fluorescence microscopy is one of the most widely used assays in biological systems. However, the technique suffers from limited multiplexing capability with previous attempts at detecting more than 11 fluorophores simultaneously resulting in barcodes that are too big for *in vivo* analysis, expensive and involve time-consuming detection schemes. Here, we introduce DNA self-assembled FRET networks that provide a unique, optical output when probed by a series of light pulses. Markov and entropy modeling of the nanoscale FRET sensors show that 125 fluorophores can be observed simultaneously. Furthermore, experimental analyses of over 1200 time-resolved fluorescence signatures show that the optical responses are repeatable 99.48% of the time and that minor variations between FRET networks can be discriminated resulting in a

total of 10^{375} unique responses. This enormous increase in spatial information density enabled by FRET networks allowed us to identify molecular signatures in lung and breast cancer tumors.

It is now known that the presence of aberrant DNA/RNA secondary structure in the regulatory regions of genes involved in cell proliferation, cells growth and apoptosis can lead to cancer. The FRET sensor we designed, self-assembles DNA probes labeled with acceptor fluorophores to the target DNA/RNA secondary structure forming an optical network. A DNA strand labeled with a donor fluorophore triplex binds to a unique sequence adjacent to the secondary structure. When the donor fluorophore is excited, the optical network results in a different optical signal based on the presence of the wild-type or the aberrant secondary structure, through which we identified lung and breast cancer cells with high specificity and over 99.9% repeatability. The small size of fluorophores results in molecular scale spatial resolution while the optical sensing mechanism enables *in vitro* and *in vivo* characterization of the structure at picosecond resolution.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my parents, grandparents and extended family for their loving upbringing, for being an inspiration, for raising us with so much comfort and so many opportunities and for giving us the flexibility to do what we want to do. Chaitan, for being a great friend!

My thesis advisor, Professor Chris Dwyer, who allowed me to work on projects that interested me, let me decide how I should solve a research problem and for ensuring that funding was never an issue while making any research related decision.

My advisor at Rice, Professor Emilia Morosan for inspiring me to be a researcher and making me believe that I could be good at research.

All of my committee members for being an inspiration and for their excellent questions and feedback and encouragement over the years.

My colleagues at Duke and Rice for making graduate school a highly intellectual and enjoyable experience.

And, all of my friends for putting this all in perspective.

1. Unclonable Cryptographic Key for Secure Authentication

1.1 Introduction

Modern cryptographic protocols rely on an asymmetry in the effort required by a legitimate user and an adversary to accurately decrypt an encoded message. These protocols ensure that communication between legitimate users is possible in polynomial time using a shared private key but a user without access to the key cannot compute the function using a probabilistic polynomial-time algorithm. In order to achieve this asymmetry, private-key cryptographic techniques currently employ algorithmic one-way functions, which are conjectured mathematical objects that are easy to compute but difficult to invert (Goldreich 2001). For the function to be easy to compute there should exist a polynomial time algorithm, f, that outputs $f(x)$ for an input x and for the function to be hard to invert, the probability of finding the inverse of an input y under f should be negligible. A sequence $\{s_n\}_{n\epsilon N}$ will be negligible in n if every polynomial p and all n's satisfy the property $s_n < \frac{1}{p(n)}$. Well-known examples of such one-way functions include the RSA and the Rabin functions. Both of these are based on the intractability of factorizing a number, which is the product of two prime numbers of comparable length. Although algorithmic one-way functions are widely used for two-factor authentication, their reliance on computational difficulty to provide security

implies that they are not protected against future advances in computational capacity or speed. For example, it was discovered recently (Vandersypen, Steffen et al. 2001) that quantum computers could run the Shor's algorithm and break RSA even though it is currently infeasible to factorize primes if each number is longer than 768 bits. Realization of quantum computers therefore, would imply that all applications using RSA for security would need to be re-evaluated. Also, use of a highly parallel network of conventional computers could potentially reverse engineer a key from the input-output pairs used in past communications. The key may also be obtained by duplicating the device. Most of the current physical embodiments of algorithmic one-way functions come with a tamper resistant packaging but remain vulnerable to sophisticated attacks. The most important practical constraint in implementing algorithmic one-way functions, however, is the difficulty and expense involved in securely embedding the required physical structure into a conventional semiconductor technology (Pappu, Recht et al. 2002).

Physical One-Way Functions were introduced a decade ago to overcome the disadvantages of algorithmic one-way functions. A Physical One-Way Function, or Physical Unclonable Function (PUF), is a physical system, which has an unknown internal state, and when perturbed by an external probe gives a unique and unpredictable output (Pappu 2001). The output of the physical structure should be reproducible yet random when perturbed by an external stimulus making it difficult to

accurately characterize or model these systems. The stimulus applied to the function is referred to as the challenge and the output is the response. It is important that the challenge not reveal any structural information about the PUF so as to maintain a significant computational advantage over an adversary. Therefore, two of the most important metrics for evaluating PUF's are the unclonability of the physical structure and the computational difficulty in accurately simulating the interaction between the probe and the physical system.

1.2. Related Work

PUF's are produced by introducing a random variation in the manufacturing process. The first of such devices constituted a transparent material doped randomly with light scattering particles or voids (Pappu, Recht et al. 2002). The device generates a speckle pattern, which is a function of the input light's wavelength, direction and depth of excitation into the doped material. The PUF, therefore, maps a unique output to each input. The speckle pattern is hashed to a binary string, which is used as a response to a challenge. Since the total number of CRP's are under 10^{69}, it is feasible to extract the structure of the PUF by applying all possible combinations of outputs or by storing all the CRP's in a look up table. Additionally, technologies do exist to extract the exact location of the doped particles in the transparent material such as invasive microscopic

sectioning or polishing, and noninvasive tomographic imaging, as described by the authors of the optical PUF (Pappu, Recht et al. 2002). Furthermore, improvements in resolution of characterization technologies since the introduction of the optical PUF, may make it possible to use techniques such as high resolution transmission electron microscopy, focused ion beam etching and nanoimprint lithography to reverse engineer the optical PUF. However, we are not aware of any attempt to commercialize or reverse-engineer an optical PUF to date. This could be in large part due to the ease of integration of electrical PUF's as opposed to the optical PUF into current CMOS devices.

The ability to produce a single unique key makes the PUF susceptible to denial of service attacks. If two or more PUF's of the same kind can be produced, the manufacturer of the PUF could retain a copy and send additional copies to legitimate users of the device. Upon arrival of the PUF, legitimate users can authenticate their PUF with the manufacturer in order to ensure that the PUF has not be replaced or tampered with. When only a single PUF of a certain kind can be manufactured, if an attacker replaces the original PUF, authentication attempts between legitimate users will fail.

Silicon PUF's have been developed and commercialized in the last few years to make unclonable RFID tags and secure FPGA's (Lee, Lim et al. 2004, Stanzione and Iannaccone 2009). In this PUF, deliberate variations are introduced in the manufacturing process, which result in random variations in the delays in the gates of the IC used as

the PUF. This generates a unique output from an IC. This output varies with the time at which the chip is probed. Therefore, a string of input bits is taken as the challenge and the response is the output of the IC at a time pre-defined by users of the PUF. Silicon PUF's have less noise compared to optical PUF's but are affected by temperature and supply voltage fluctuations. Random variations in the manufacturing process have also been exploited to make ring-oscillator PUF's where random variations in the manufacturing process produced a PUF with a unique frequency response to an applied signal (Suh and Devadas 2007). Arbiter PUF's contain k delay stages and two inputs, which are made to race through the k delay stages. Based on which signal reaches an "arbiter" element in the last stage first, a zero or one is assigned to the output. Unfortunately such an architecture is susceptible to machine learning attacks. In order to increase the strength of the PUF, an XOR Arbiter PUF was designed, where there are a few Arbiter PUF's operate in parallel and the output of all the PUF's is XORed and the final output determined. Similar to the XOR-Arbiter PUF, a Lightweight Secure PUF was designed with multiple Arbiter PUF's operating in parallel but with each PUF having a different input.

Magnetic PUF's consist of a slurry of magnetic particles of different shapes and sizes mixed together randomly (D.G. Porter and Muller 1994). Every PUF has a distinct magnetic signal that is amplified by a magnetic head. Even though the device is highly complex, the output suffers from a lot of noise. More importantly, this device may not

even strictly fall under the class of PUF's since every device produces a fixed output regardless of the input. There are no challenge-response pairs involved in authentication, which makes it very easy for an adversary to imitate the response even without gaining physical access to the device. It does have applications in credit cards for instance where every user has a specific physical key, which is matched to their credentials every time a card is swiped.

Coating PUF's were developed shortly afterwards and used an IC on which wires are deposited in the form of a comb structure which is subsequently coated with a material doped randomly with dielectric particles (B. Skoric and Tuyls 2006). The challenge is a voltage of a particular frequency and amplitude and the response is a bit string of capacitance values at specific points in the sensor network. This PUF has the advantage of being inseparably bound to the underlying IC which makes it a good control PUF. Using this technique, the authors claim to be able to get 100 bits/mm^2 of information from the coating reproducibly, which is a relatively large area by IC standards.

While the PUF's mentioned above require modifications to the manufacturing process or require a new device to be engineered, PUF's based on SRAM are intrinsic to the electronic device using the SRAM (Guajardo, Kumar et al. 2007). The authors identified that each SRAM reproducibly outputs a voltage within 10 ns of it being

powered on because of static leakage currents that arise due to the weak coupling between the two inverters in the SRAM. Since they use only the characteristics during start-up, it is possible to get PUF behavior and ensure that the circuit performs the desired functionality at the same time. Different SRAM's produce different outputs, therefore a string of responses from different SRAM's is defined as a challenge and the unique output voltage obtained is the response. The SRAM based PUF suffers from a low number of CRPs and hence is a weak PUF.

1.3. Authentication Schemes

In this section, we compare PUF's with the traditional, more frequently used authentication schemes, such as password based schemes and smart cards.

The password-based scheme is currently the most frequently used authentication scheme owing to its ease of use. However, it is also the weakest form of authentication. In this scheme, each user is assigned a username and a secretly determined unique string of characters or a password. The user typically memorizes the password since storing it may result in easy access to the password by an adversary. As a result of this, passwords are short, relatively simple and commonly tied to some credentials of the user such as names or birthdays, which makes the scheme susceptible to brute force

attacks. Alternate ways to extract the password from the user include eavesdropping or social engineering attacks. The most serious concern with the password-based scheme is the passive nature of an attack. After an adversary obtains a user's password, he has access to the user's information but the user is unable to detect an intrusion if their content remains unmodified.

One-time passwords overcome some of the limitations of traditional passwords. A one-time password can only be used once even by legitimate users. In such a scenario, even if the adversary gains access to the user's password, they can only access the user's current information but not to any future modifications to the information. However, the passwords still remain vulnerable to the brute force or social engineering attacks. Importantly, the user does not detect an intrusion in the system while the adversary gains access to all of the user's current information.

Hardware based keys such as smart cards provide higher security than password based schemes because of their relatively hard to alter characteristics. Smart cards make use of a physical key in addition to a password or PIN and are thus classified under two-factor authentication schemes. The PIN is known only to the user and is used to access the smart card. Information stored on the card, such as the user's private key, is sent directly to the computer to authenticate a user (MicrosoftCorporation 2013). Smart cards are also designed such that all the security related operations are performed in the

processor on the card and the software on the host computer cannot observe these operations. Additionally, since a single copy of the card exists with the user, the prospect of a missing card being quickly noticed is high. It is therefore a stronger authentication scheme than the password-based scheme since the attacker needs physical access to the user's smart card. Additionally, a brute force attack to obtain the private key may not work but a brute force attack to retrieve the PIN may be possible. Unfortunately, this additional security comes with added material and support costs. Traditional smart cards are expensive to purchase (both cards and readers must be supplied to employees), and they can be easily misplaced, stolen or cloned. PUF's have the advantages of the smart card over passwords and in addition, they are physically unclonable and have a large number of challenge response pairs, which means that an adversary cannot forge a single signature as with a smart card.

Despite the advantages of PUF's over traditional authentication schemes, the security of the PUF's described above are based on computational difficulty and are not strictly unclonable. All of these PUF's achieve 'unclonability' by making use of an uncontrollable aspect in the manufacturing process in order to realize a single device that has a low probability of being similar to other devices manufactured by the same process. Unfortunately such an approach can be prone to sophisticated attacks through which the underlying structure of the PUF, and hence the message being exchanged can be obtained. In (Ulrich, 2010), the authors created an accurate software model of an

arbiter PUF and Ring-Oscillator PUF's of any size and XOR Arbiter PUF, Feed-Forward Arbiter PUF, Lightweight Secure PUF up to a certain size and complexity by applying machine learning techniques. Their model was able to reproduce the CRP's of the original PUF in most cases. The training time varied from a few seconds in the case of the Arbiter PUF to months in the case of the Lightweight Secure PUF. The authors of this work suggest that an Arbiter PUF with 8 XOR's and 512 bitlength is secure and beyond the reach of modeling attacks. However, XOR Arbiter and Lightweight Secure PUF's of up to 16 XORs and up to 512 bitlength were reverse-engineered using modeling attacks in conjunction with power and timing side channel attacks in (Ruhrmair, Xu et al. 2013). Unfortunately, increasing the number of XOR's in a PUF, decreases the stability of the PUF's response (Herder, Yu et al. 2014). Therefore, currently, the number of XOR's cannot be increased arbitrarily in order to make the PUF resilient against modeling attacks. The SRAM PUF has also been cloned by authors in (Clemens and Christian 2013). The PUF was first characterized using Photonic emission analysis, where the number of photons emitted by the p-channel and n-channel transistors during their on and off state is examined in order to extract the contents of the embedded memories. Subsequently, focused ion beam trenching was employed to re-create the PUF. In principle, this technique could be employed to clone other CMOS based PUF's since they rely on the infeasibility of characterizing gate delays in IC's. A significant drawback of existing PUF's is the ability of an adversary to introduce data and time triggers. A

data trigger is a special instruction or data that could affect the normal operations of the chip or introduce additional operations. A time trigger is designed to turn on an alternate path in the circuit at a pre-determined time. This is especially problematic since data and time triggers could be designed to go unnoticed during testing of a chip.

A key limitation of existing PUF's is that only a single unique device can be manufactured. This necessitates that the PUF be transferred between communicating parties each time information is to be exchanged, in order to avoid the need for a trust anchor, which greatly slows down the communication process and increases the risk of the PUF being stolen. Alternatively, a large number of CRP's of each PUF can be stored on a server through which a user can authenticate their device. However, the chief purpose of introducing randomness in the manufacturing process is to ensure that the need for a trust anchor, in this case the manufacturer of the PUF, is eliminated. If the PUF associated with each user and the CRP's used to authenticate each PUF are stored on a server, this server would have to serve as a trust anchor. Furthermore, the main purpose of using a PUF for authentication is that security from a physical key is more robust than that derived from software approaches. By storing all the CRP's used to authenticate a PUF on a server, security is now shifted from the physical key to the attacker's ability to access the data stored on the server.

In this report, we propose a new cryptographic key that has the advantages of a PUF but is truly unclonable, identical copies of the key can be created at the time of manufacture and the response of the key varies with time.

2. Resonance Energy Transfer based key

We introduce a Resonance Energy Transfer (RET) based key, which is truly unclonable even when the adversary gains unfettered physical access to the key. The RET-key is a physical system constituting a DNA grid on which fluorophores are attached at arbitrary positions. The probe is multiple pulses of light at various wavelengths projected on the grid at different relative time delays. The interaction between the input light and network of fluorophores results in resonance energy transfer between the fluorophores, which begins an exciton mixing process within the fluorophore network. The extent of RET and exciton mixing results in time-resolved fluorescence emission from the key that follows a multi-exponential decay. This decay has a characteristic intensity, offset and set of lifetimes that collectively serve as the response from the RET-key to the initial challenge. Therefore, the challenge constitutes a series of light pulses of specific wavelength and delay and the response is a time-resolved histogram of the resulting fluorescence decay.

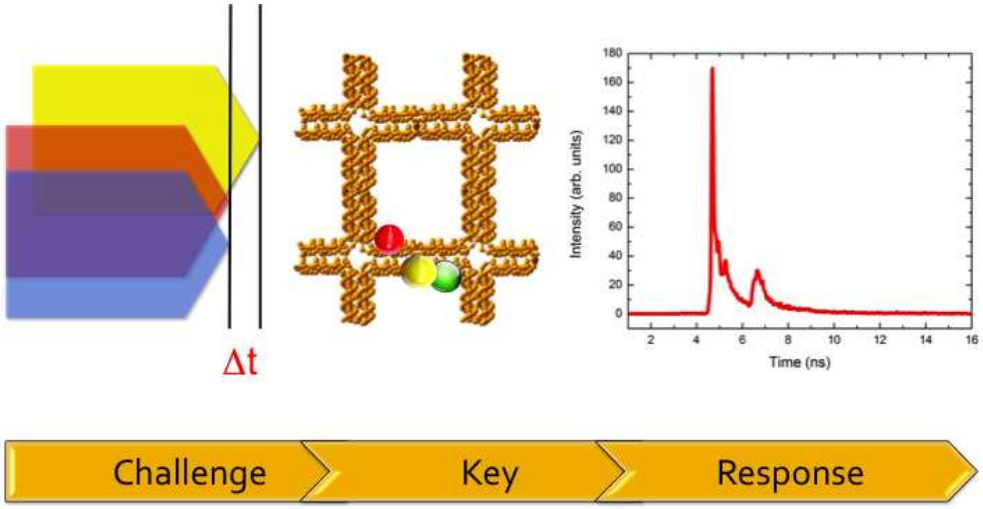
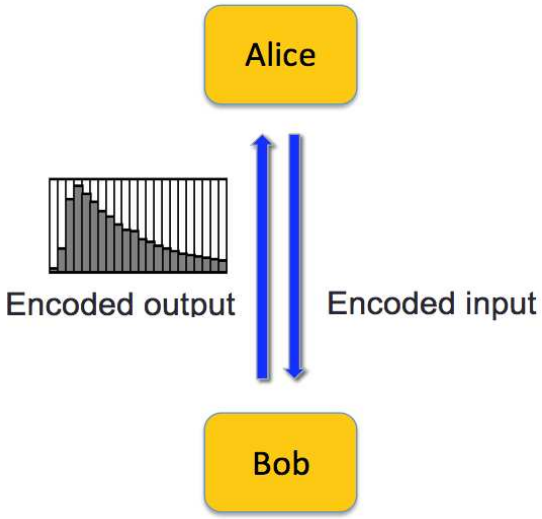


Figure 1: Authentication using the resonance energy transfer based key involves applying a series of light pulses of specific wavelengths and delays to a fluorophore network placed on a nanoscale DNA grid. The output is a time-resolved fluorescence histogram that is unique to the combination of excitation conditions and key used.

Scheme 1 shows the process of verifying with the RET-key that an authorized transmitter, using an established protocol, sent a message and that the message has not been modified or tampered with. If Alice wants to authenticate Bob, she sends a challenge constituting the excitation wavelength and delay to him. In addition to the challenge, a set of helper data is transmitted to minimize the difference in the measurements between the two parties. Bob then computes a histogram by applying the challenge on his key and sends the output to Alice. Since Alice and Bob possess identical keys, Alice checks the histogram from Bob against her own and if both sets of data match within the allowed tolerances, authenticity is established.



Scheme 1. Example of an authentication scheme using our device. Alice sends a obscured-challenge constituting the excitation wavelength and delay to Bob. Bob then computes a histogram by applying the challenge on his key and sends the output back to Alice. Since they possess identical keys, if the histograms obtained by Alice and Bob match, authentication is established.

The challenge and response strings are in the public domain but they are XORed with a cryptographic obscurer derived from the key. The obscurer is derived from the output bit string obtained independently by Alice and Bob when they apply an identical, public input to their respective keys. The large output space enabled by our scheme ensures that an adversary will have a 1 in 10^{375} chance of randomly identifying the correct cryptographic obscurer. Without the correct obscurer, even if an adversary steals a key, they cannot ensure authentication. Making use of a cryptographic obscurer,

therefore, strengthens our protocol against passive attacks (Stinson 2006). A more detailed authentication protocol can be found in Chapter 9.

The strength of the RET-key depends on the assertion that small changes to the network, input wavelength or excitation time produces a significant change in the output. This results in a large number of possible configurations making it infeasible for an adversary to simulate the RET-key and therefore has a negligible probability of arriving at the right response. True unclonability can be achieved because there is no technology available currently to detect the precise location of these fluorophores since they are spaced much closer than the diffraction limit of light as will be detailed later in the text. Moreover, we are able to make two identical keys, which will eliminate the need to physically transfer the device between the two communicating parties or store the CRP's on a server. The RET-key thus has the potential to overcome the deficiencies present in existing keys.

2.1 RET-key Fabrication and Characterization

The nanoscale key structures are constructed using hierarchical DNA self-assembly. We program individual DNA strands to self-assemble into cruciform shaped structures called tiles (Figure 2 (c)) that have sticky ends which enable different tiles to self-

assemble into a grid structure as shown in Figure 2(a) (Pistol, Mao et al. 2010). DNA self-assembly has several advantages over other fabrication techniques to create RET networks:

1. Functional groups can be attached as close as 0.34 nm.
2. DNA can be programmed to self-assemble into a wide variety of structures with predictable local geometries.
3. DNA self-assembly has well-developed synthesis chemistry and the enzymes, reagents are readily available.
4. Long DNA sequences are commercially available at relatively low cost. The estimated cost of any given RET-key is <$0.01.
5. The self-assembled structure is rigid.
6. The structure has extremely high stability and longevity when stored under the right conditions.
7. DNA self-assembly can be highly parallel. We can make ≈10^{12} DNA grids in only a few hours' time. The grids are at a concentration of 250 nM and the total volume is only 60 μl.

Each DNA grid is composed of 16 tiles and is 80x80 nm in size. In a DNA tile, each core contains 100 bases, the four shells contain 168 bases and the four arms contain 124 bases, which result in 392 bases per tile. We fabricated DNA grids with 16 tiles, which

result in 6272 bases per DNA grid. The grid is fully addressable and fluorophores may be placed as close as 0.34 nm to form a RET network. We purchase HPLC purified DNA strands with fluorophores pre-conjugated at the 3′ or 5′ ends from IDT-DNA. We make use of an Atomic Force Microscope (AFM) to characterize and determine the yield of our structures. AFM images were obtained on an Agilent PicoLE equipped with OTR-8 tips (Veeco). Figure 2(b) shows the relatively large number of intact 16 tile DNA grids formed by the process described above. Capillary gel electrophoresis is used to give us information regarding the yield of the conjugation.

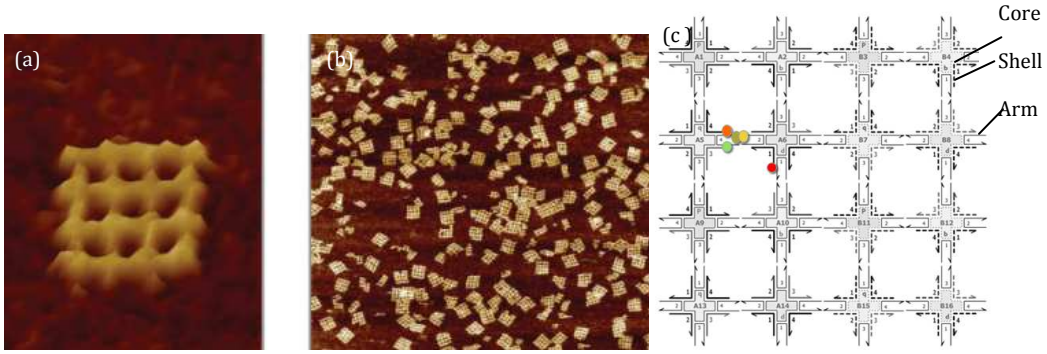


Figure 2: AFM image of a 16 tile DNA grid (Pistol, Mao et al. 2010) (b) AFM image indicating good yield of the self-assembled DNA structures. (c) Layout of a 16 tile grid indicating core, shell and arm strands.

Resonance energy transfer occurs between two fluorophores when the emission spectrum of the donor fluorophore overlaps with the absorption spectrum of the

acceptor fluorophore. Dipole-dipole interactions between the donor and the acceptor result in the transfer of energy between the two molecules. The rate of energy transfer from the donor to the acceptor is given by:

$$k_T(r) = \frac{1}{\tau_D}\left(\frac{Ro}{r}\right)^6$$

where Ro is the Förster radius which is determined by the extent of spectral overlap, r is the distance between the donor and the acceptor and τ_D is the lifetime of the donor in the absence of energy transfer. Lifetime indicates the time it takes for 63% of the fluorophores to decay to the ground state from the excited state. The extent of RET determines the intensity and time between the absorption of light and the subsequent reemission which are the two observables of interest for the key. Several processes such as complex formation, solvent effects, translational and rotational diffusion, wavelength and input time of the excitation pulse can modulate both the output intensity and lifetime. The difficulty in independently estimating the value for each of these parameters for every fluorophore in the network has made it infeasible to precisely model well-mixed RET networks when the number of fluorophores is even modestly large (e.g., >3) (Watrob, Pan et al. 2003). This problem is compounded by the fact that the values of these parameters are not constant even for a specific fluorophore and change based on a number of factors including the excitation conditions, microenvironment of the RET network and the interaction network of each fluorophore. Remarkably however,

when multiple batches of the same RET key are fabricated and measured independently, the optical responses are highly reproducible, as will be shown in Chapter 8.

2.2. Unclonable Keys

The manufacturer of the RET-key can create $\approx 10^{12}$ identical DNA grids in one run of the manufacturing process. Depending on the characterization technique used, the total number of DNA grids is orders of magnitude more than that required for a single RET-key. The DNA grids obtained from a single run of the manufacturing process can therefore be split into several identical RET-keys, which in turn can be distributed between multiple users of the key. We assume at all times that an attacker has complete knowledge about the manufacturing process, the DNA sequences being used in the keys, the reagents used for self-assembly and the type of fluorophores compatible with DNA. Thus, if an attacker can decipher the underlying fluorophore network in a particular RET-key, they will be able to physically re-create any number of identical keys. However, the RET-key and the protocol constructed for its use ensures that it is infeasible to reverse-engineer the underlying fluorophore network in a particular RET-key. The primary method used to characterize DNA nanostructures is AFM while the primary method to determine fluorophore position currently is fluorescence microscopy and super resolution imaging. We will show in Chapter 8 that the poor lateral resolution of AFM makes it infeasible to decipher atomic scale features in the RET-key's structure.

The high level of precision and control over the location of the fluorophores in a RET-key results in current microscopy and imaging techniques being unable to extract the precise location of the fluorophores if they are spaced a few nanometers from each other. In order to resolve two adjacent points using light microscopy, they have to be separated at least by λ / 2[NA] where lambda is the excitation wavelength and NA is the numerical aperture of the objective. Objectives usually come with a NA less than 1.5 and the minimum wavelength in the visible is around 400 nm, which puts the lowest lateral resolution achievable at around 250 nm. This number is significant since the fluorophores in the RET-key can be placed as close as 0.34 nm making it infeasible for any of the current conventional microscopy techniques to detect the precise location of the fluorophores on the grid. Fluorescence spectroscopy can be used to determine the number of fluorophores of each type on a key but the locations of the fluorophores and hence the interaction network of the key cannot be determined. Super resolution techniques such as stimulated emission depletion microscopy may be used to resolve fluorophores separated by over 5nm but this has been shown only on individual fluorophores (Rittweger, Han et al. 2009). These techniques involve spatially or temporally modulating the state of the fluorophores by selectively quenching fluorescence in a specific sequence. After this, the size of the point-spread function of the emitting fluorophore is minimized and its center calculated. In our key, we have clusters of fluorophores and are able to place 13 fluorophores in a 5 nm x 5 nm area. This is

assuming that the fluorophores can be placed at every 4th base without compromising the yield of the self-assembled structure. It is therefore not possible to detect a shift in emission after quenching resulting in the position of the fluorophores being undetermined. Fluorescence imaging with one nanometer accuracy (FIONA) may be used along with STED to improve the image resolution. FIONA fits a two-dimensional Gaussian to resolve the position of a dye molecule with 1nm precision but this has been demonstrated only for a single dye (Yildiz, Forkey et al. 2003).

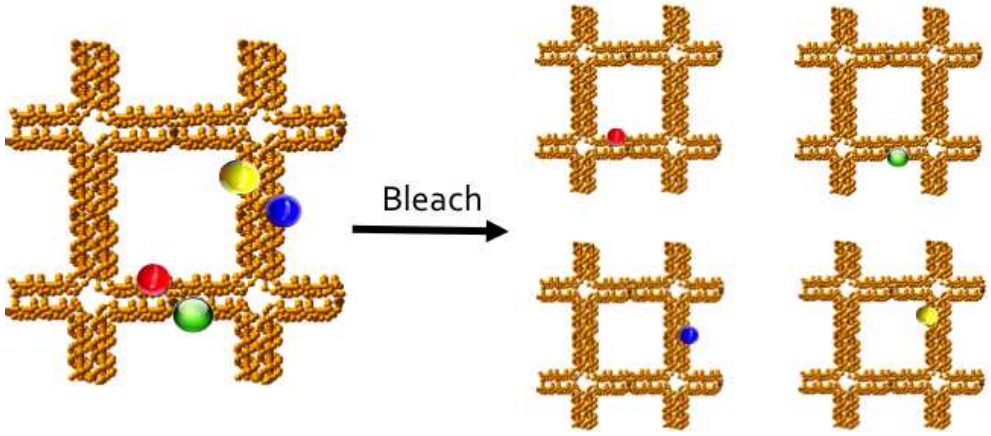


Figure 3: The ideal way for an attacker to use super-resolution imaging to determine the position of the fluorophores is to bleach the RET network until a single fluorophore remains on the grid. An adversary may then use super-resolution combined with FIONA to obtain position information.

The lateral resolution of other super resolution techniques such as 3-dimensional stochastic optical reconstruction microscopy (3D STORM), Photo-activated localization microscopy (PALM) and Near-field Scanning Optical Microscopy (NSOM) appear to be

inferior to that of STED for clusters of fluorophores (Gordon, Ha et al. 2004, Rust, Bates et al. 2006, Huang, Wang et al. 2008, Shtengel, Galbraith et al. 2009). 3D STORM can achieve axial resolutions of 50 nm, however, the DNA grids measure only 2 nm axially. In order to use STORM/PALM an attacker would have to bleach all but one fluorophore in the network in order to obtain the position information of the fluorophore as shown in Figure 3. While such an attack can determine the position of the fluorophore that is most red in wavelength, exciting the remaining fluorophores would result in RET and therefore will not result in the target fluorophore being bleached. Anti-dye and anti-hapten antibodies may be used to chemically bleach the target fluorophore but they are only available for select fluorophores.

In order to protect against all super-resolution attacks, we propose the use of dark quenchers as shown in Figure 4. Dark quenchers behave similar to regular fluorophores, in that, excitons would transfer to the dark quencher based on the distance and orientation of the quencher with respect to the fluorophore. Quenchers are available that quench the fluorescence of all fluorophores while others quench the fluorescence in certain wavelength bands. One major difference between a fluorophore and a dark quencher is that any exciton that gets transferred to a dark quencher, will not be re-absorbed by the network or fluoresce. It will instead leave the network as heat. It is therefore not possible for super-resolution techniques to detect the position of the

quencher. However, the RET network and hence the signature of the key will change significantly due to the presence of quenchers. When a pulsed light source is used to excite a fluorophore, the heat emitted by the nearby quenchers generates an acoustic wave in the medium. This technique known as photoacoustic microscopy can be used to measure the distance between a FRET pair by measuring the output acoustic energy. This technique will not work with the RET networks because of 4 reasons: 1) The location of the fluorophore being excited should be known in order to estimate the position of the quencher. 2) This technique has only been demonstrated on a single FRET pair. When the interaction network becomes more complex, exciting a single fluorophore without partially exciting other fluorophores becomes difficult. Moreover, every fluorophore may have multiple quenching paths, which can make modeling the fluorophore-quencher interaction challenging since the underlying network is unknown. 3) Bleaching all the fluorophores and quenchers in the network except for one pair could work along with a super resolution technique. But as mentioned above, because the RET networks are closely coupled, bleaching the intended fluorophore-quencher pair is not possible due to RET. The most blue pair may be observed but not the remaining fluorophores and quenchers. 4) Modeling the network can be made more difficult by having only black quenchers in the network. Black quenchers quench the fluorescence of all fluorophores irrespective of the emission wavelength. In such a scenario, trying to bleach all the fluorophores and only analyze the quencher network will not work, since

we get a collective signature from all the quenchers. Any attempt to bleach a single quencher will bleach all the quenchers in the network.

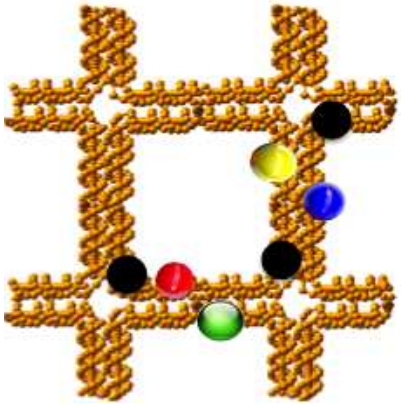


Figure 4: The black circles indicate dark quenchers placed on the DNA grid. The dark quenchers change the RET network similar to regular fluorophores, however, they do not fluoresce making it impossible for super-resolution techniques to detect their position.

Non-optical characterization techniques to determine the fluorophore structure such as ultracentrifugation (strand separation by weight) or DNA sequencing would require the disintegration of the DNA grid and individual analysis of each DNA strand. Such an approach will not result in the characterization of fluorophore type and position on the DNA strands because of three reasons: First, there are multiple fluorophores on each strand, therefore obtaining weight information will not give us the right fluorophore combination or the precise location of a fluorophore on the strand, both of which are extremely important in determining the output of the key for a specific challenge. Second, let us assume that there is only one fluorophore on each strand. Since

all the DNA tiles share the same core and shell sequences as shown in Figure 5, these strands and hence the fluorophores on them cannot be traced back to their respective tile. Third, once the structure of the grid is lost, it cannot be re-annealed to the original structure since the core and shell strands would be misplaced from their original location. Alternatively, restriction enzymes can be used to cleave the specific sequence of DNA at the intersection between two tiles, without disrupting the entire structure. However, methlytransfereases may be used to protect duplex DNA from restriction enzymes. An attacker, therefore, cannot separate out individual tiles and reverse-engineer the location of a specific fluorophore within the RET network.

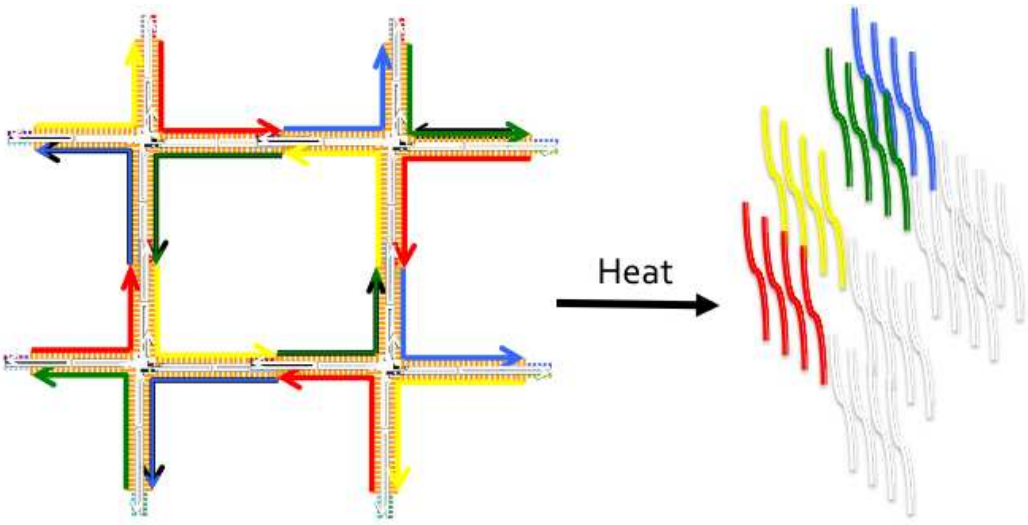


Figure 5: Denaturation of the DNA grid to analyze individual strands of DNA and hence determine the RET network will not work since all the tiles in the grid share the same core and shell sequences. We could determine the position of the fluorophores on each strand but the strand cannot be traced back to its original tile.

Other non-optical characterization techniques such as scanning probe methods to determine the underlying RET network currently lack the resolution to identify the location of fluorophores within a network since they may be placed as close as 0.34 nm from each other. To prevent future advancements in scanning probe characterization techniques, a 3-dimensional RET key can be fabricated with the fluorophores placed within the 3-dimensional structure. DNA self-assembly offers an advantage in this regard since 3-dimensional DNA nanostructures of different sizes and shapes have been fabricated and any attempt to disrupt the outer layers of the self-assembled structure will result in the loss of the entire DNA structure and hence the RET-network itself

(Han, Pal et al. 2011). As described above, due to the hierarchical nature of self-assembly, once the DNA structure is lost, it cannot be re-annealed to its original configuration. There are several electron microscopy techniques that can currently reach sub-nm spatial resolutions. In order to obtain high spatial resolution, however, electron microscopy typically requires a conducting sample. The DNA and the fluorophores used in the RET-keys are electrically non-conducting. Transmission electron microscopy on unlabeled DNA nanostructures was recently attempted, however, the resolution was worse than that of conventionally used AFM techniques (Buckhout-White, Robinson et al. 2013). Raman spectroscopy has not yet been used to study DNA nanostructures. Characterizing even duplex and single stranded DNA using Raman spectroscopy can be challenging due to molecular conformation or packing density of the DNA molecules on a metallic substrate, which is required for Raman spectroscopy. Thermal cycling pretreatment (heat the DNA to 95 °C and then cool in an ice bath) of the DNA may be used to overcome the molecular conformation or packing density problem (Barhoumi and Halas 2010), however, thermal cycling of the RET networks will result in the disintegration of the DNA structure and the irreversible loss of the original RET network.

For the reasons described above, we would ideally like to place fluorophores at every base but this could potentially interfere with the assembly of our grids. It is

extremely important that our structures are well formed since the relative distance of the fluorophores with respect to each other determines the output of the key and having poorly assembled grids will result in deviations from the expected network and output of the key.

In order to maximize the number of sites that can be conjugated on our grid, an additional structure, which we call a 'DNA bridge' was designed. The bridge is essentially a DNA duplex which bridges the cavities in the grid as shown in the Figure 6. A 5 base DNA sequence projects out from tiles 6 and 7, which enables the bridge to bind to the center cavity in the grid. The bridge has DNA sequences at its ends that are complementary to those at tiles 6 and 7. Therefore, once the 16-tile grid is assembled and the DNA bridge introduced into the solution, the bridge self-assembles with the sequences at tiles 6 and 7 across the center cavity of the grid. The bridge, being 2 nm wide, is too small to probe using AFM. We therefore attach a vitamin called biotin to the center of the bridge and introduce a protein called streptavidin in to the solution after the bridge binds to the grid. Streptavidin has a high affinity for biotin, and is 6.8 nm across its long axis, at a pH of 7.5 (Weber, Ohlendorf et al. 1989). This makes it easily discernible with the AFM tip we use (nominal size = 15 nm). As shown in the AFM images, Figure 6(a) is a control sample of 16 tile grids with no bridge attached. In Figure 6(b), streptavidin was added to the 16 tile grid-bridge structures and annealed at 4° C for 2 hours. The presence of streptavidin can be seen as white spots in the AFM image

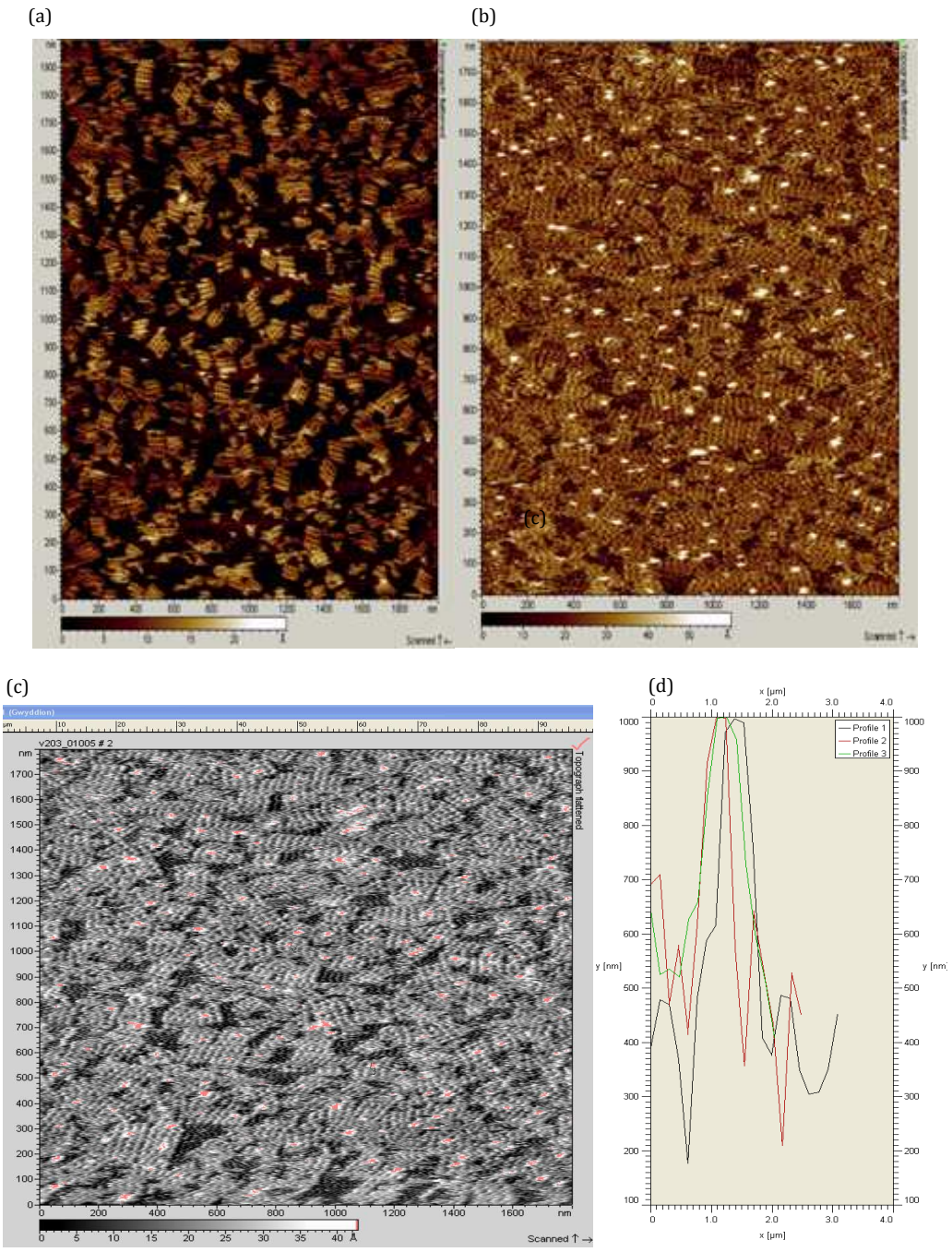


Figure 6: (a). AFM image of control sample consisting of 16-tile grid with no bridge added. (b) AFM image of streptavidin-biotin laden bridge structure demonstrating white spots in the center cavity indicating increased z height of the streptavidin molecules. (c) Gwiddeon image calculating the precise center of the streptavidin molecules. (d) Line profile across 3 bridge structures showing similar peak positions indicating that the bridge is rigidly bound on both sides to the grid.

because of its elevated height compared to the grid. AFM images tend to increase the lateral dimensions of streptavidin and hence the software Gwyddeon was used to identify the precise location of streptavidin as shown in Figure 6(c). Figure 6(d) shows the axial profile across 3 center cavities where streptavidin was noticed. They do align within nanometers of each other indicating that the bridge was bound on both sides of the cavity. If the bridge bound to only one side, it would have been displaced by the AFM tip during imaging resulting in the streptavidin center being shifted significantly. The bridge is 52 bases long and can be placed in all of the 9 cavities in a 16-tile grid, which leads to 468 additional sites on the grid at which fluorophores can be placed.

3. Continuous-time Markov Model of the RET-Key Structures

There are two important factors that need to be taken into account when evaluating the RET-key: (1) Repeatable signatures for the same key under identical excitation conditions and (2) Any change of the input parameters (wavelength, excitation time) or the key configurations (fluorophore type and position), should result in a unique signature (amplitude and lifetime). In theory, varying the excitation wavelength in a multi-fluorophore key should result in different fluorophores being excited, which results in different pathways for resonance energy transfer, and in turn changes the amplitude and lifetime of the donor and acceptor. Varying the excitation time will result in an offset in the histogram, which can be used as an additional parameter in the signature. A change in fluorophore will result in a change in the output since there will be a change in the value of R_o and κ^2, as described in chapter 2. A change in the position of the fluorophore will result in a change in the value of r in the Förster equation, which again should result in a different signature. In an attempt to discretize the values of wavelength, excitation time and r for which we obtain a unique signature, we simulated the response of our keys using a Markov model by changing each of these values individually.

3.1. Continuous-time Markov Model

In probability theory, stochastic processes are defined as a set of random variables indexed in time (Trivedi 2001). The values that the random variable can take are known as states and there exists a predefined probability at which transitions can occur between these states. If the probability of transition varies with time, the process in termed non-homogenous and if the transition rate is independent of time, it is referred to as a homogenous process. The state space along with time can take continuous or a discrete set of values. A stochastic process is referred to as a Markov process if the transition to the next state is only determined by the current state and does not depend on the behavior of the system in previous states, that is, it is a memoryless system. The time that a homogenous, continuous time Markov chain spends in a given state is known to exhibit an exponential behavior similar to that of the time spent in the excited state of a fluorophore (Trivedi 2001). We model the RET-key structures as a continuous time, discrete state Markov chains. The fluorophores each constitute a random variable and are defined as being in state 1 if they are excited and state 0 if they are in the ground state. The probability of transition between states is given by the transfer rate in the Förster equation. In order to model fluorescence, we make use of an absorbing state or a state from which no output transitions are allowed. This implies that once the exciton enters the absorbing state corresponding to a specific fluorophore, we

assume it has left the system or fluoresced with the emission characteristics of that fluorophore.

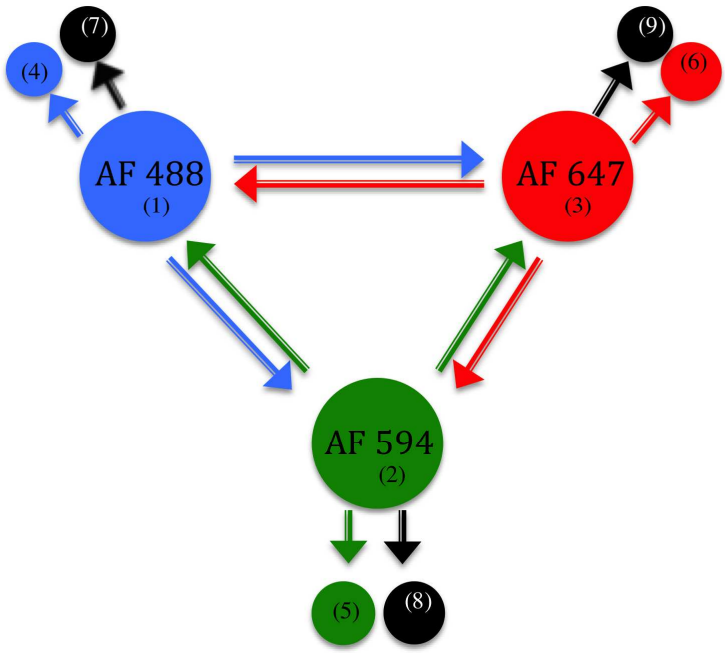


Figure 7: Model of a 3-fluorophore key with AF 488, AF 594 and AF 647 as the fluorophores. The smaller circles denote the absorbing states of the corresponding fluorophores indicated in the larger circle. The black arrows indicate the loss of the exciton through non-radiative decay.

In the example shown above, we choose three fluorophores AF 488, AF 594 and AF 647. The three smaller circles represent the absorbing states of these fluorophores. The transfer between these fluorophores can be written using the Kolgomorov differential equations:

$$\frac{d\pi j(t)}{dx} = \sum_{i \neq j} \pi i(t) aij - \pi j(t) aj$$

Here denotes the state of the system and , are the transition rates between states. Again referring to the example in Figure 7, the transitions between the various states may be written as follows:

—(1) = -(a12+a13+a14+a17)*x(1) + a21*x(2) + a31*x(3)
—(2) = a12*x(1) - (a21+a23+a25+a28)*x(2) + a32*x(3)
—(3) = a13*x(1) + a23*x(2)- (a31+a32+a36+a39)*x(3)
—(4) = a14*x(1)
—(5) = a25*x(2)
—(6) = a36*x(3)
—(7) = a17*x(1)
—(8) = a28*x(2)
—(9) = a39*x(3)

Once the state probability is known, the probability density function may be used to determine fluorescence from each state. We used this continuous-time Markov model to generate decay curves for various excitation-key combinations and thus determine if a significant change is observed in the amplitude and lifetime values for different fluorophore networks and excitation wavelength combinations. The implementation of the algorithm may be found in Appendix A.

3.2. Variation in signature with fluorophore separations

In the first experiment, we used the Markov model to generate histograms for various fluorophore separations. We modeled a two-fluorophore system consisting of AF 488 and AF 594 separated by 1, 4, 10 and 20 nm. In all cases, AF 488 was directly excited and fluorescence from AF 594 recorded. As can be seen in Figure 8, significant changes in the output are observed even for small changes in the distance. When AF 488 and AF 594 are placed at a distance of 1 nm and AF 488 is excited, it is expected that AF 488 acts as the donor and AF 594 acts as the acceptor for RET. This should result in a decrease in AF 488's intensity and increase AF 594's intensity. Further increase in distance should increase AF 488's intensity and reduce AF 594's intensity since the amount of energy transfer is inversely proportional to sixth power of the distance between the two fluorophores. We notice some energy transfer from the donor to the acceptor close to $2R_0$ (11.89 nm) but the transfer is negligible at a separation of 20 nm.

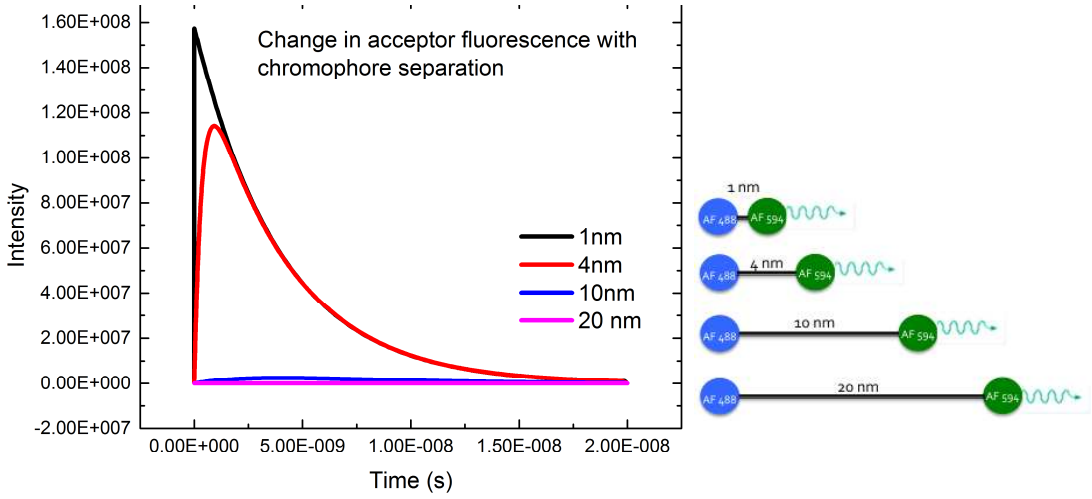


Figure 8: Markov model results indicating changes in histogram with change in fluorophore separation. In all four cases, AF 488 is excited and the fluorescence from AF 594 is observed.

3.3. Variation in signature with excitation wavelength

In an attempt to observe a change in signature on changing the excitation wavelength, we modeled a two-fluorophore system consisting of AF 488 and AF 594 at a fixed distance of 4 nm. Changing the initial conditions of the fluorophores modified the excitation wavelength. For instance, ignoring extinction coefficient and quantum yields, we notice in Figure 9 that at around 495 nm, AF 594 is only 10% excited while AF 488 is 90% excited, the initial conditions to model 500 nm would be (0.9, 0.1) for AF 488 and AF 594 respectively. When only AF 488 is excited the excitation condition is denoted by (1,0)

and when only AF 594 is excited, the excitation condition is denoted by (0,1). The output histogram shows a significant change in all three cases as shown in Figure 10.

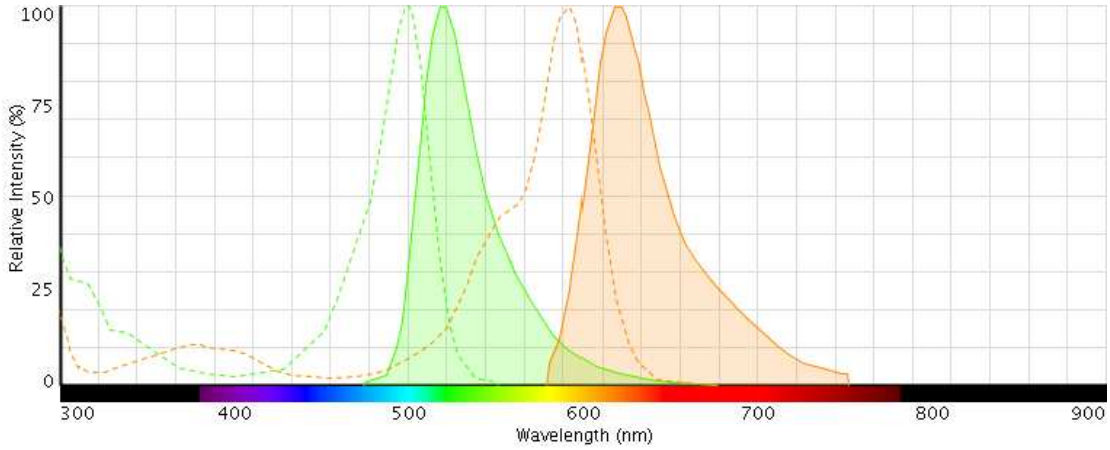


Figure 9: The excitation (dashed green) and emission (solid green) spectra of AF 488 and the overlap between the excitation (dashed orange) and emission (solid orange) of AF 594 (LifeTechnologies 2013)

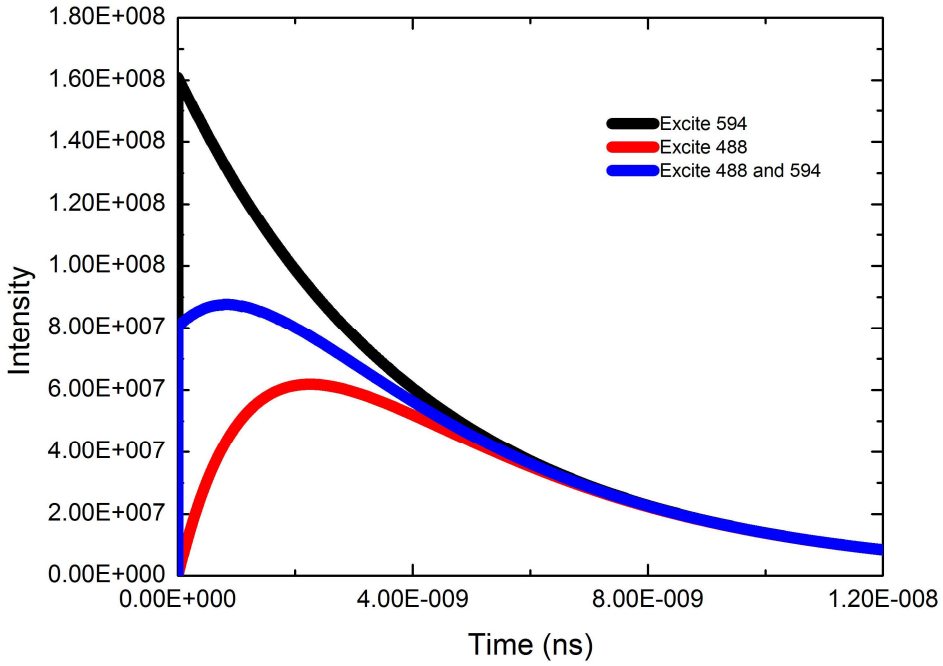


Figure 10: Markov model results indicating changes in histogram with change in excitation wavelength.

3.4. Variation in signature with number of fluorophores

We observed the change in fluorescence of the donor AF 488 on increasing the number of acceptors. When there is only a single acceptor in the system (AF 594), the black curve in Figure 11 shows the fluorescence of AF 488. However, on adding one more acceptor (AF 647) to the system, energy is transferred from AF 488 to both AF 594 and AF 647 resulting in lower fluorescence from AF 488 represented by the red curve in Figure 11.

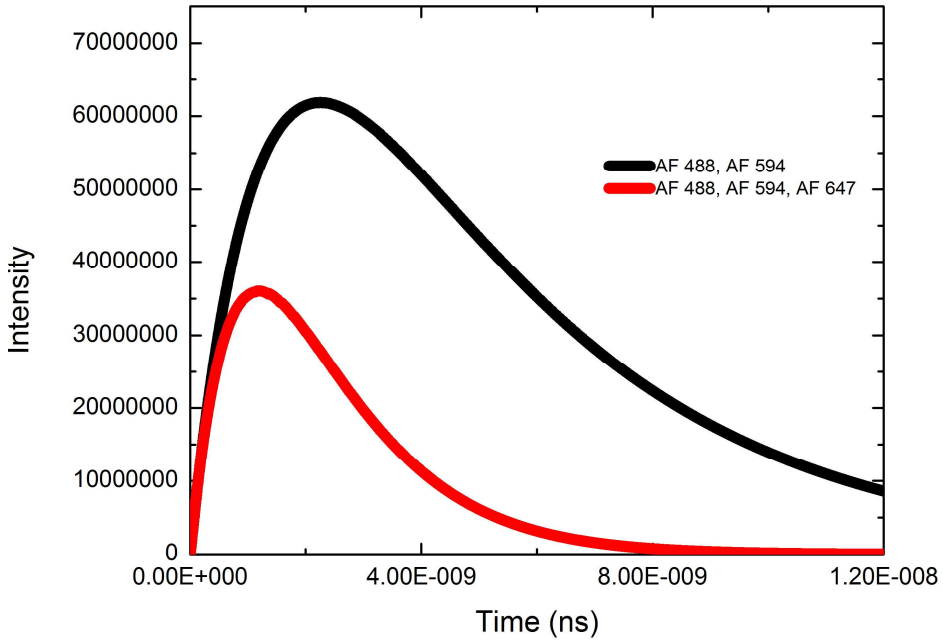


Figure 11: Markov model results indicating changes in histogram with addition of fluorophores.

In conclusion, the continuous time Markov model results were consistent with the behavior predicted by Förster's equation. We noticed significant variations in the key signatures on changing the position of a fluorophore by 1 nm, by modulating the excitation wavelength or by the addition of a single fluorophore to the network. In the following chapter, we will describe a technique that can be used to experimentally generate time-resolved fluorescence decay curves from RET networks.

4. Time-Correlated Single Photon Counting

In order to generate signatures for the RET-keys experimentally, we can either use steady state or time-resolved fluorescence signals. Steady state fluorescence measurements limit the total entropy that can be achieved using our keys since present day instrumentation can only detect 11 fluorescent bands uniquely. Therefore, we use time-resolved fluorescence to compute the signatures of the RET-keys. Owing to the fast decay time of the fluorophores that we use in our keys (100's of ps), we are unable to apply single-shot analog recording techniques to capture the time-resolved fluorescence signal. For instance, we need to collect data points at least every 100 ps if the fluorescence decay of a fluorophore is 1 ns. This is not possible using present-day transient recorders (Lakowicz 1999). Additionally, since we operate at extremely low sample concentrations, we lose many photons due to the collection optics, detector sensitivity, etc. This results in very few photon counts per 100 ps time bin, which is insufficient to get a high signal to noise time-resolved fluorescence decay. Increasing the laser power in order to obtain more counts is not possible since fluorophores have a tendency to photo-bleach at high laser intensities, which will further reduce the photon counts. Since current transient recording devices lack the capability to collect photons at the resolution needed, we decided to use single-photon time-resolved fluorescence decays as signatures of our keys.

There are several techniques to measure time-resolved intensity decays including time-gated detection, streak cameras and Time-Correlated Single Photon Counting (TCSPC). Time gated detection suffers from low sensitivity and a high degree of systematic errors compared to TCSPC. Streak cameras are considerably faster and provide more resolution than the other two techniques. Additionally, they enable simultaneous detection of both wavelength and time decays, which could significantly increase the total number of resolvable signatures from our keys. However, they suffer from low dynamic range of measurable intensities, rarely exceeding 1000, and have poorer signal to noise ratio compared to TCSPC (Lakowicz 1999). These two factors in turn will result in a lower number of resolvable signatures for our keys because of the limited number of resolvable amplitudes and the inability of the system to resolve finely separated complex decays. Also, TCSPC boards are significantly less expensive than commercially available streak cameras. We, therefore, use TCSPC to measure the time-resolved fluorescence decays from RET-keys.

TCSPC uses periodically generated light pulses to capture single-photon fluorescence events from the sample over many cycles. The collected single-photon events are later binned with a resolution of 10's of ps and a time-resolved fluorescence histogram is generated, as shown in Figure 12. This resolution is usually determined by the timing jitter introduced by the excitation source, the detector and the electronics used in TCSPC.

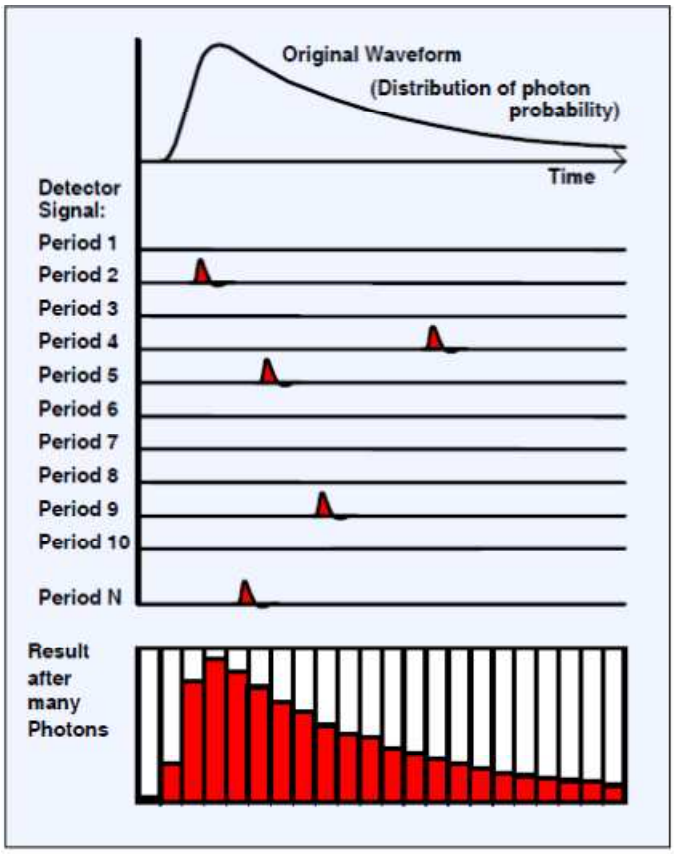


Figure 12: Time correlated single photon counting builds a histogram of counts versus time channels to create the fluorescence decay curve (Becker 2008).

The principle of operation of time-correlation single photon counting is described in Figure 13. The laser releases a short pulse of many photons periodically, which is used as a start signal to trigger a counter (here a TAC with a charge fraction discriminator). The first photon detected by a single photon detector (here a single photon avalanche photodiode), generates a stop signal and turns off the counter. The time difference between the start and stop signal is recorded and is used to increment

the value of an appropriate time bin in a fluorescence histogram as shown in Figure 12. After accumulation of arrival times from a large number of photons, the fluorescence histogram can be read out.

We use a bh SPC – 130 module for TCSPC. The module has a Constant Fraction Discriminators (CFDs), Time-to-Amplitude Converter (TAC), a fast Analog-to-Digital Converter (ADC), data processing logic, and memory integrated on the board as shown in Figure 13. A more detailed description (derived from the fourth edition of Becker and Hickl GmbH's TCSPC handbook) of TCSPC with respect to this module is given below:

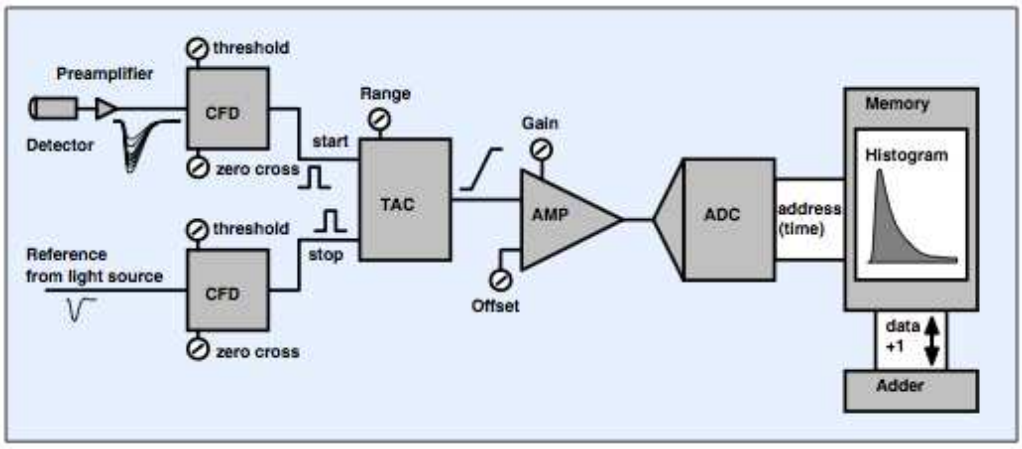


Figure 13: Architecture of TCSPC in reversed start-stop mode (Becker 2008).

The signals generated by the reference light source and detector are first sent to their respective constant fraction discriminators (CFD's) to extract precise timing information. A simple leading edge detector will not suffice for our purposes due to

varying pulse amplitudes. The CFD overcomes this problem by detecting a constant fraction of the pulse amplitude. Comparing the original signal with an amplified, inverted and delayed version of it avoids pulse-height induced timing jitters. The CFD and SYNC parameters are specific to the set-up used. We use the following CFD and SYNC parameters to generate the TCSPC signatures:

CFD:

Limit Low: 0, ZC Level: -10.58

SYNC:

Freq Div: 1, Threshold: 0, ZC Level: -12.09

The output pulse of the CFD from the laser and the detector are used as the start and stop pulses respectively of a time-to-amplitude converter (TAC). The TAC then generates an electrical signal proportional to the time difference between the laser and the detector pulses. The start pulse is used to charge a capacitor while a stop pulse is used to stop charging it. The voltage accumulated by the capacitor in the start-stop interval determines the time of arrival of a photon at the detector with respect to the laser pulse. Time differences of a few ps can be resolved using this technique. We use the following TAC settings to generate our time-resolved histogram:

Offset: 1.96, Gain: 1, Range: 5×10^{-8}, Limit High: 97.65, Limit Low: 3.53

The output from the TAC is then sent to a biased amplifier where the gain and offset of the voltage signal may be adjusted. The amplifier compresses the full-scale conversion range of the TAC to a smaller time window. The signal is then sent to a fast flash analog to digital converter (ADC) where the voltage signal is converted to a digital photon detection time with respect to the laser pulse. The output of the ADC is used to increment the memory location that corresponds to the digital address of the photon detection time. The ADC should accurately resolve the TAC signal into the right time channel with good linearity across the entire time range. Failure to do so will result in noise in the channels and potentially distort the histogram. Furthermore, it should be extremely fast in order to shorten the dead time of the system. We use the following settings to obtain a high-resolution signal at high speed:

Operation mode: FIFO, Time/channel: 1.22×10^{-11} s, Time/Div: 5×10^{-9} s

Collection time: 100 s, Repetition time: 500 s, Display time: 0.1s, Time per point (ms): 10

In order to process incoming data and display histogram memory in parallel, the TCSPC system switches between two memory blocks. The time-resolved histogram is built up over time by incrementing the contents of the memory locations corresponding to the detection time of the photons.

The time-resolved fluorescence decay generated by TCSPC is unique to the combination of the RET-key and the excitation wavelength and delay used. TCSPC is

often used on a small number of fluorophores to determine the distance between the fluorophores, the viscosity of the solution, ion concentration, the amount of oxygen, local refractive index, pH, fluorophore aggregation and protein binding (Lakowicz 1999). The above-mentioned properties are most often measured in biological systems for studying biochemical interactions, detection of abnormalities as well as to characterize drug delivery. There are no instances, to the best of our knowledge, where the time-response from a large collection of fluorophores has been used for biological sensing. The ability to get different biological species to form a RET network and examining the time response of the fluorescence from the network could result in a parallel, sensitive and cost-effective sensor. This avenue is explored further in Chapter 11.

In the Chapter 5, I will describe a set-up that will enable us to probe our keys with multiple excitation beams and record the time-resolved fluorescence signature of the keys.

5. Time-Resolved Excitation Optical Set-Up

We previously showed that varying the excitation conditions, that is excitation wavelength and delay, result in unique signatures for the same key. In an attempt to observe and quantitate the expected theoretical behavior with experiments, we custom-built an optical Time Resolved Excitation (TREX) setup shown in the schematic below.

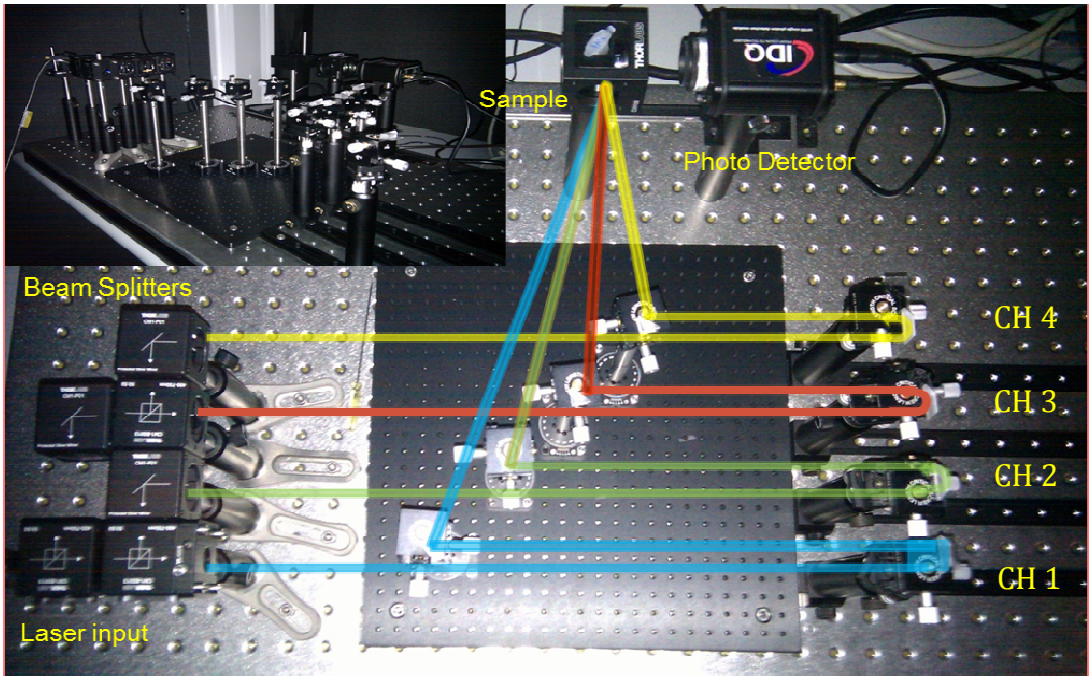


Figure 14: The key is probed with a challenge using this optical set-up. The incoming white light is split into 4 parts with varying wavelengths and delays. The beams excite the sample whose fluorescence decay is used as the response. The inset shows the schematic's side view.

The instrument uses a picosecond NKT photonics SuperK Extreme supercontinuum pulsed laser with a repetition rate of 80 MHz and maximum power of 3W. The input laser power is split equally into 4 parts using Thorlabs Pellicle beam splitters. Each of the 4 parts functions as an independent excitation channel with adjustable wavelength and time delay.

For wavelength selection, we considered using monochromators and bandpass/laser line filters. Monochromators facilitate higher signal to noise ratio owing to adjustable slit widths but at the expense of light intensity. Additionally, there are several sources of unwanted light in monochromators. Imperfect gratings may result in stray light transmission and ghost images and higher order diffraction at the grating may result in the excitation light passing through the exit slit when the emission wavelength is a multiple of the excitation wavelength. Monochromators also introduce wavelength dependent time delays and/or broaden light pulses. Bandpass/laser line filters are usually used to overcome the above limitations. A wide collection of effective and inexpensive laser line filters are available. We use Thorlabs UV/Vis laser line filters since a large collection of wavelengths are available with a FWHM as low as +/-1.2 nm. The transmission in the blocking region is as low as <0.01% (OD>=4) in these filters. Neutral density filters may be employed to attenuate light equally at all wavelengths. The optical density of neutral density filters can be varied in order to match the intensity

of light coming from the 4 light beams on TREX. Wavelengths can be selected ranging from 350 nm to 700 nm.

Each beam travels through a dovetail prism and is redirected to the sample using a silvered right angle prism. The distance through which the beam travels and hence the time at which it is incident on the sample is determined by the location of the prisms. The dovetail prisms are mounted on post-holders, which slide along rails on an optical breadboard. The delay on each beam can be set from 0-4ns by moving the post-holders on the optical rails. A fine-delay option is available for each channel using translational stages located under each dovetail prism. Moving the translational stages enables the user to adjust the delay with a precision of 10 ps. We notice that moving the dovetail prism by 1 inch introduces a delay of 0.169 ns as shown in Figure 15. However, this delay changes with the wavelength of light used as shown in Figure 16. We measured a scattering sample at a fixed position of the dovetail prism on TREX and noticed different delays on changing the excitation filters as shown in Figure 16. Table 1 lists the available excitation delays on the four channels of TREX.

The 4 beams are then deflected onto the sample using mirrors. The sample is 50 µl of a fluorescently labeled DNA tetramer at a concentration of 250 nM as described in Chapter 2. The sample volume and concentration are chosen to obtain high photon counts for the time-resolved signatures and to prevent inner-filter effects that may result

from very high concentrations (Lakowicz 1999). High sample concentrations could also result in the excitation light being absorbed by the surface facing the light source, which in turn results in low photon counts at the detector.

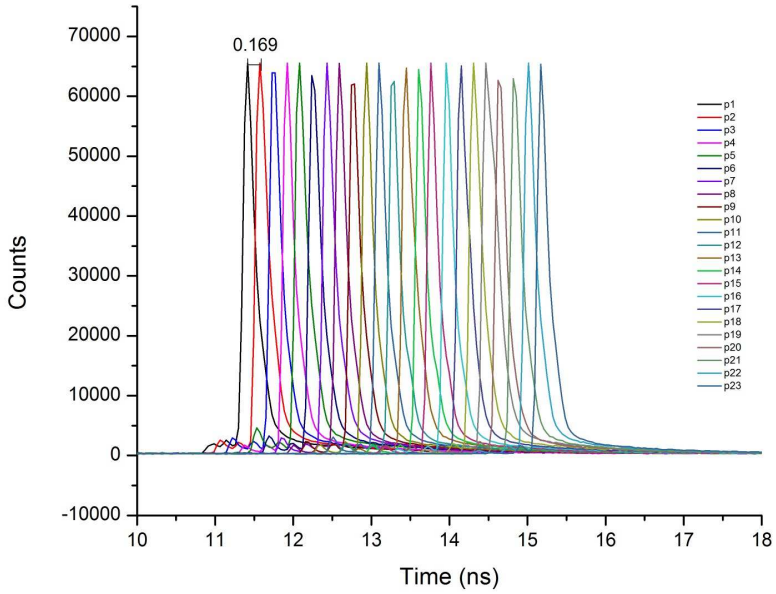


Figure 15: The figure shows that an incremental delay of 0.169 ns is introduced on moving the dovetail prism in increments of 1 inch.

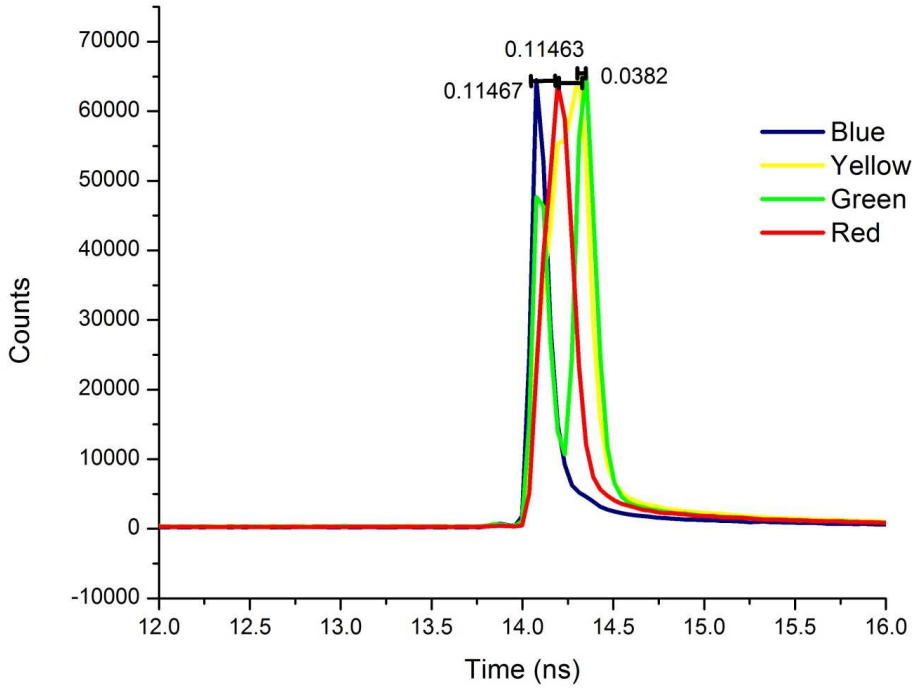


Figure 16: Change in delay with wavelength

Table 1: Range of excitation delays on the 4 channels of TREX

Position	Ch 4	Ch 3	Ch 2	Ch 1
1	1.29	1.4859	1.5468	1.9018
2	1.46	1.6559	1.7168	2.0718
3	1.64	1.8359	1.8968	2.2518
4	1.82	2.0159	2.0768	2.4318
5	1.96	2.1559	2.2168	2.5718

6	2.14	2.3359	2.3968	2.7518
7	2.31	2.5059	2.5668	2.9218
8	2.49	2.6859	2.7468	3.1018
9	2.68	2.8759	2.9368	3.2918
10	2.82	3.0159	3.0768	3.4318
11	2.99	3.1859	3.2468	3.6018
12	3.17	3.3659	3.4268	3.7818
13	3.33	3.5259	3.5868	3.9418
14	3.51	3.7059	3.7668	4.1218
15	3.67	3.8659	3.9268	4.2818
16	3.84	4.0359	4.0968	4.4518
17	4.033	4.2289	4.2898	4.6448
18	4.21	4.4059	4.4668	4.8218
19	4.34	4.5359	4.5968	4.9518
20	4.55	4.7459	4.8068	5.1618
21	4.72	4.9159	4.9768	5.3318
22	4.91	5.1059	5.1668	5.5218
23	5.06	5.2559	5.3168	5.6718

For the detector, we considered using Photo multiplier tubes (PMT's) and single photon avalanche photo diode (SPAD's) as they are frequently used to detect low levels of light. SPAD's are a better choice than PMT's due to their high speed, low cost and high gain. However due to the smaller active area diameter of a SPAD (20 μm) compared to the active area of a PMT (≈1 cm^2), a substantial amount of signal may be lost. This disadvantage may be overcome by highly focusing the beam using focusing lenses as shown in Figure 17. Also, SPAD's traditionally have a high timing FWHM resulting from the wavelength dependent tail that follows each pulse. The SPAD that we use, however, has a maximum timing FWHM of only 60 ps.

Fluorescence from the sample is observed using the SPAD at right angle to the direction of the incident light as shown in Figure 14 in order to avoid light from the laser being incident directly on the detector. The signal observed at the SPAD is transformed into a 4096-bin histogram using TCSPC. The resulting histogram is a convolution of the instrument response function (IRF) and the measured fluorescence decay. The IRF is the response of the instrument to a zero-lifetime sample or a delta function and can be classically interpreted as the narrowest pulse that can be measured by an instrument. The IRF varies with the instrument used and the wavelength. On the TREX, the measured IRF values are listed in Table 2.

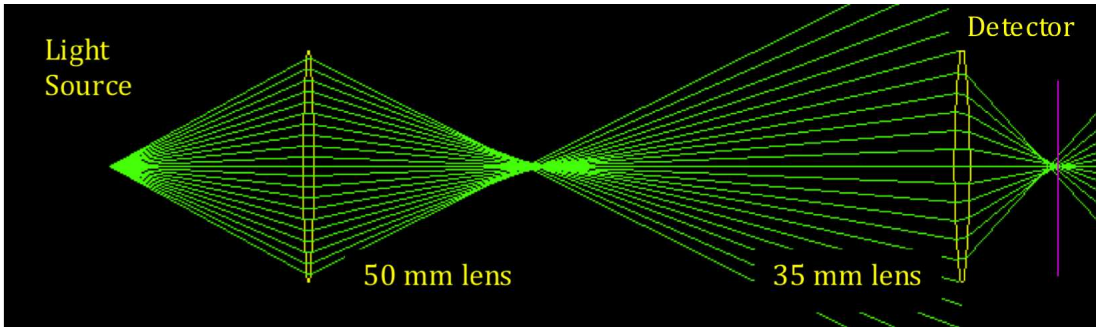


Figure 17: Lens arrangement between the sample and the detector to focus the fluorescence from the sample onto the SPAD (http://www.ub.edu/javaoptics/index-en.html). The above arrangement resulted in 3 orders increase in fluorescence, significantly improving the signal to noise ratio and hence achieving near ideal intra-key and inter-key correlation.

Table 2: Variation of the TREX Instrument Response Function with wavelength.

Wavelength (nm)	IRF (ps)
488	127.61
543.4	122.85
620	111.19
670	110.56

It is possible to deconvolve the IRF from the measured fluorescence decay, if sufficient photon counts are available to make a continuous signal time approximation, but it is not required for our measurements. The IRF chiefly depends on the instrument

and the wavelength of light being used, which remain unchanged between measurements. Hence, we chose not to deconvolve the IRF while obtaining our time-resolved signature. We instead use time-gating to eliminate source contribution to a signature. We discuss the generation of time-resolved fluorescence signatures using TREX in Chapter 7.

It is worth noting here, that TREX can serve as a powerful tool for biological assays. The most important requirements for instruments studying biological systems are high sensitivity, single photon detection capability, high time resolution, multi-wavelength capability, high speed and depth measurements (Becker 2008). TREX has all of the above properties and should be able to perform optical sectioning when a high power picosecond laser, that can enable two-photon excitation, is employed. In addition to the above properties, the variable time delays of the excitation pulse can provide additional opportunities for a large range of time-resolved signatures, which in turn, can enable sensitive detection of multiple biological processes in parallel.

TREX is the first instrument that can generate signatures of RET networks when probed with multiple excitation wavelengths and delays. However, it suffers from a large bench footprint, relatively high cost and a manual alignment process that can introduce errors between the measurements of the communicating parties. The signal-processing techniques in TREX use analog circuits, which consume a large area and

make the integration of large arrays of detectors impractical. To create a digital, high-speed detector with a small area, we propose a novel, parallel read-out CMOS design for the detection of a SPAD signal. In order to achieve this, we design a circuit for time to digital conversion, which can be used to process the SPAD's signal efficiently. The design is implemented in 0.5 μm CMOS technology. The circuit operates at 370 MHz, which allows for a time resolution of 2.7ns. However, our design can be scaled to a newer technology in order to increase the time resolution.

6. Parallel, High Speed, Digital Signal Processing Design for Optical Detectors

The work in this chapter was done under Professor James Morizio in collaboration with Rajat Chada, Sergej Deutsch and Kai Hu for the Digitial CMOS Design Methodologies class at Duke University.

Systems-on-Chip with a combination of photonics and electronics are increasingly prevalent today because of their high power efficiency and speed. The high-speed signals obtained with photonic chips require precision detectors with the capability to detect and process single photons. Single Photon Avalanche Photodiodes (SPADs) currently provide the highest timing accuracy and sensitivity compared to other optical detectors and are capable of single photon resolution. The output of the SPAD is readout using several analog elements, such as high-speed analog switches, high-precision current sources, time-to-digital converters, and analog-to-digital converters. A conventional SPAD is manufactured in a standard low-cost CMOS technology.

In order to take advantage of the high sensitivity provided by SPADs, we need appropriately fast techniques to process the signal generated by the SPAD. Processing

refers to deciphering what information the photon contains and readout refers to organizing and transporting the signal either directly from the SPAD or some derivative of the signal. There are several different processing techniques such as time-uncorrelated, time-correlated, and spatio-temporal correlated techniques. In particular, we are interested in time-correlated techniques since we want to capture information regarding the arrival time of the photons. The timing information of the detected photons is collectively analyzed in order to obtain the fluorescence signal of the sample.

The current readout and processing techniques we use, do not exploit the high sensitivity and speeds that may be obtained using SPADs. In addition, these processing circuits use analog elements, which require a large area and therefore make the integration of large array of detectors impractical. We propose a novel readout and processing architecture that provides a solution to the problems mentioned above. Our novel design makes use of digital elements, which significantly reduces the area requirements. Additionally, the parallel architecture enables processing of every detected photon, thereby lowering the total integration time.

6.1. High-Level Description

The high-level goal of our project is to obtain timing information of every photon that is detected. In order to achieve this, we measure the time difference between the rising edge of the reference signal and the detected signal. A histogram of the arrival

times of all registered photons is generated and this is taken as the fluorescent signal emitted from the sample being measured.

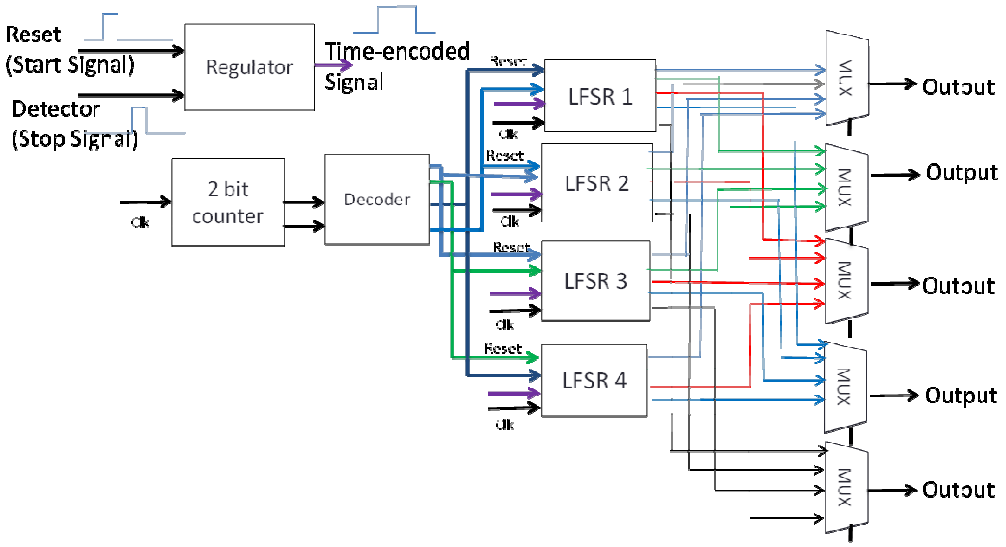


Figure 18: High level block diagram of the readout and processing architecture

Table 3: Synchronization signals for the LFSR and the MUX

	Read	Count	Reset
LSFR	1	2	3
LSFR	2	3	4
LSFR	3	4	1
LSFR	4	1	2

The block diagram of our readout and processing architecture is shown in Figure 18. The output of the detector is analyzed using multiple counting elements in parallel, which necessitates the need for synchronization signals. We use two clock signals in our implementation. A faster clock CLK1 is used for counting and a slower clock CLK2 is used for synchronization.

A two-bit counter is used as a clock divider and generates the outputs 00, 01, 10 and 11. These signals are given as input to the decoder, which converts them into a one hot representation. The regulator generates a pulse which is high for the period between the start and stop signals. This pulse serves as the enable signal for the counter. When the enable signal is high, the counter is operational and when the enable signal is low, the counter holds its previous value. In our circuit, we implement the counter using a Linear Feedback Shift Register (LFSR). The number of counting elements required is determined by the time it takes to transfer the timing information off chip. We assume that it takes 20 ns to shift the signal out; however, we get a signal out of the SPAD every 5 ns thus necessitating the use of 4 LFSR's. The 4 LFSRs used are synchronized to count, read and reset according to Table 3.

A reset signal first sets the value of the LFSR to a predetermined value. In the next clock cycle, the enable signal is set to high in order to enable that LFSR to count. In

the following clock cycle, the value of the LFSR is read. These three operations are distributed among the various LFSRs in order to ensure parallel operation. That is, when LFSR 1 is being read, LFSR 2 is counting and LFSR 3 is reset so that it can count in the next clock cycle. The output of the LFSR is 6 bits long. These 6 bits serve as inputs to 6 MUX's from which the final count is read out. A detailed description of the function and implementation of all the blocks follows.

6.2. Regulator

We designed a regulator to measure the time difference between the generation of the laser pulse and the detection of the optical signal. The regulator produces a time-encoded pulse, with width equal to the working period of the LSFR counter. The output of regulator is connected to the ENABLE input of the LFSR. The LFSR will therefore work only when regulator output is high. When the regulator output is low, the LFSR will keep the counting result or get reset.

Figure 19 shows the schematic of the regulator, implemented using an SR latch. When "En" is high, the regulator keeps its value. When "Set" is low, the latch is set to "1" and when "Reset" is low, the output is forced to "0". The LFSR can therefore be activated only when the counter is counting (Table 3). In the top-level diagram (Figure 18), we see that the "Set" and "Reset" signals from the regulator are connected to the

reset of LFSR and stop signal, respectively. During the falling edge of the reset, the regulator output will go high, which enables the LFSR to count. With the arrival of the stop signal, the regulator output will go low and the counting stops.

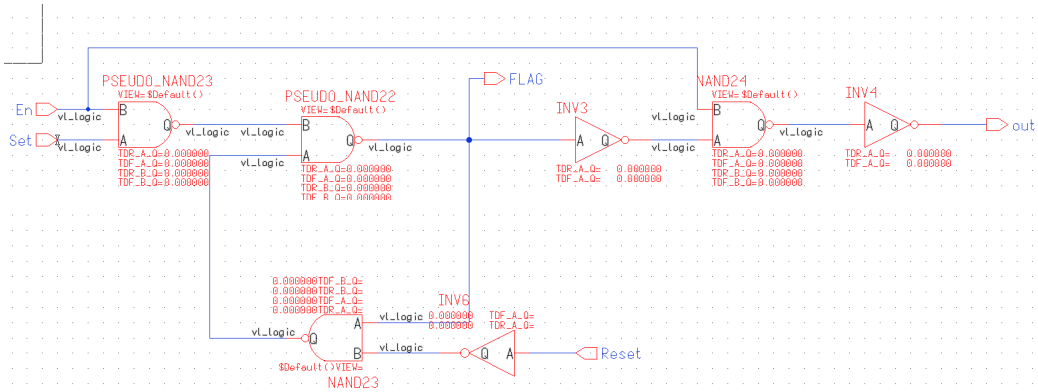


Figure 19: Schematic of the Regulator

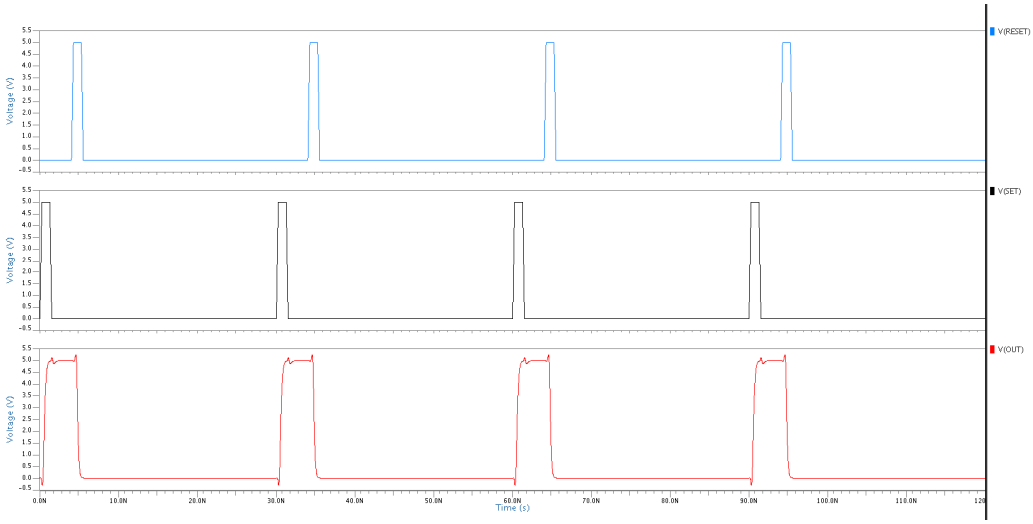


Figure 20: Analog simulation of the regulator.

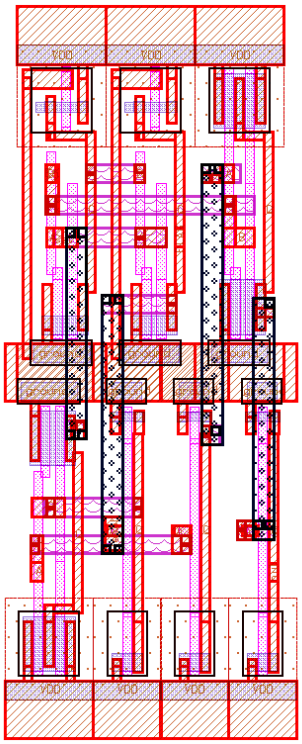


Figure 21: Layout of the regulator.

6.3. Counter

The counter is implemented by connecting a series of D flip-flops. A total of 2n outputs are possible when n flip-flops are used. The D flip flop is implemented with preset and clear functionality as shown in Figure 22. This was done in order to enable circuit operation independent of the clock, when required. For the counter, we set the preset and clear high such that the output follows the input on the rising edge of the clock pulse and holds that value until the next rising edge is detected.

Table 4: Timing information for the D flip-flop with clear and preset.

INPUT				OUTPUT	
CLK	CLR	PRE	D	Q(t+1)	Q_bar(t+1)
X	0	1	X	0	1
X	1	0	X	1	0
X	0	0	X	1	1
1	1	1	X	Q(t)	Q_bar(t)
0	1	1	X	Q(t)	Q_bar(t)
↑	1	1	0	0	1
↑	1	1	1	1	0

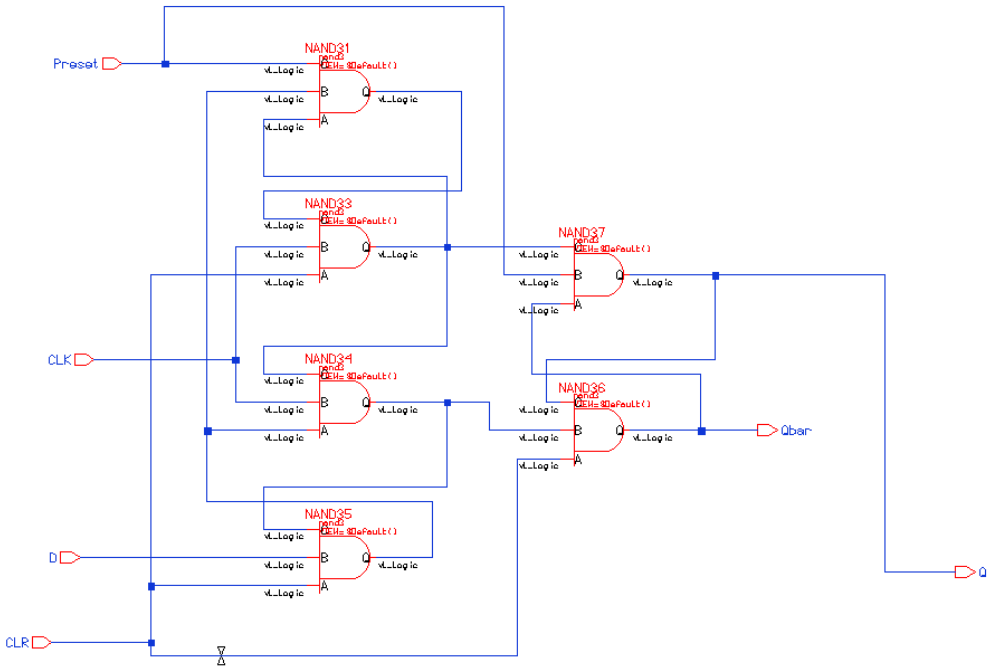


Figure 22: Schematic of the D flip-flop.

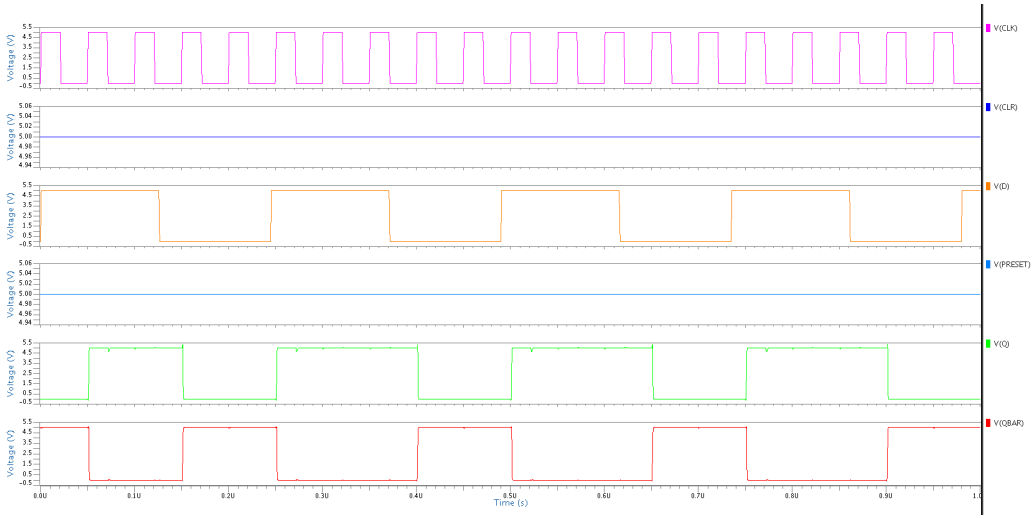


Figure 23: Analog simulation of the D flip-flop.

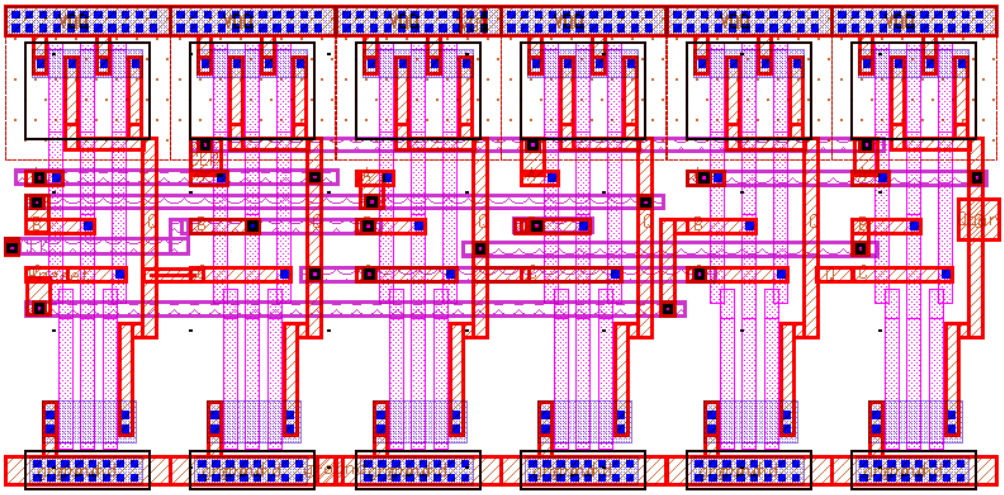


Figure 24: Layout of the D flip-flop.

We see that a single clock signal serves all the four flip-flops in the counter. The Q output of each flip-flop is the D input of the next flip-flop. The Q output of the fourth flip-flop is the D input of the first flip-flop. A single Clear signal is connected to the Clear inputs of the first three flip-flops, and the Preset input of the fourth. Therefore, when the Clear signal is 1, then Q_0, Q_1, and Q_2 are all 0 and Q_3 is 1. When Clear goes low, the Clock causes the outputs of the flip-flops to change as shown in Table 4. First Q_0 is 1, then Q_1 is 1, followed by Q_2, and Q_3 and the cycle repeats.

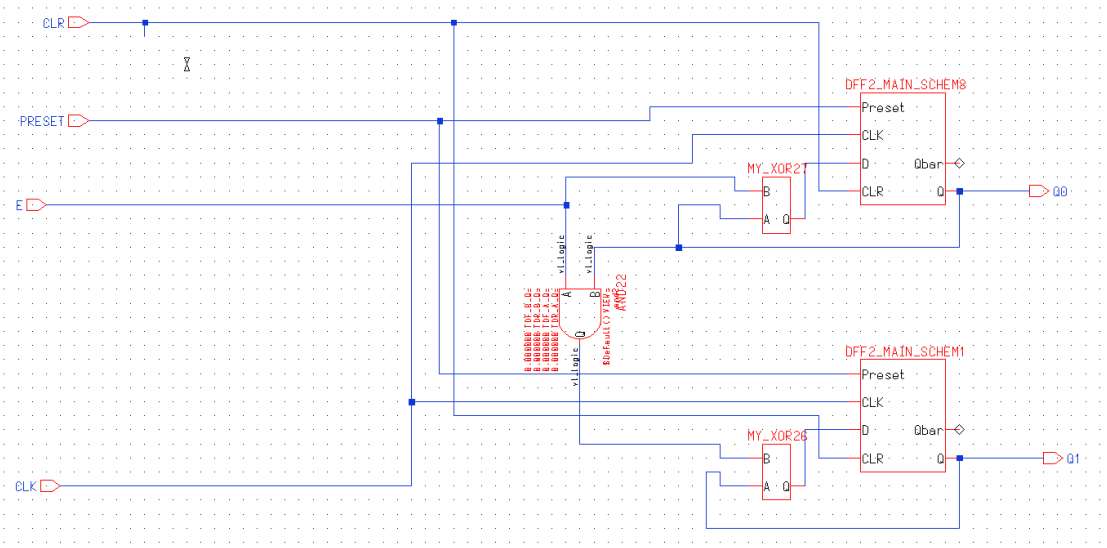


Figure 25: Schematic of the 2-bit counter

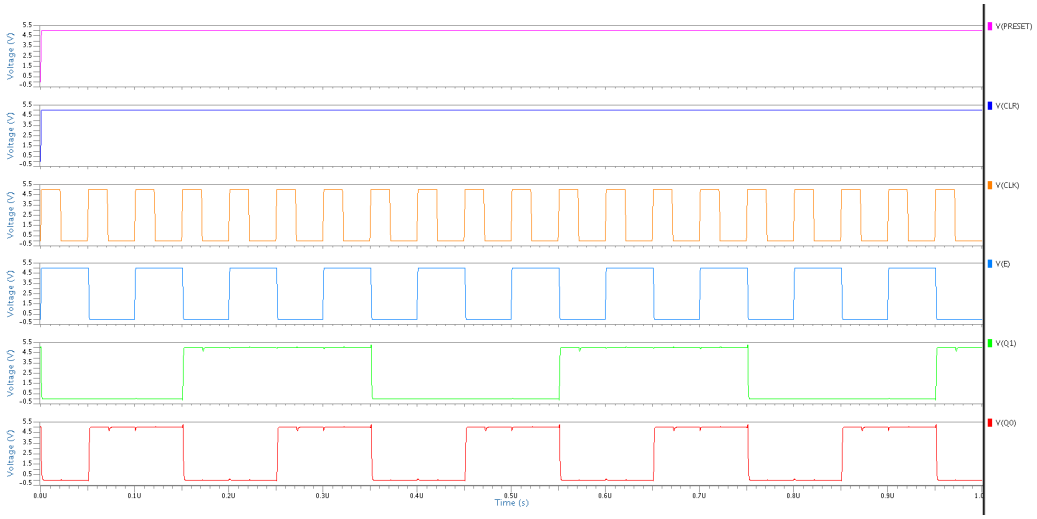


Figure 26: Analog simulation of the 2 bit- counter

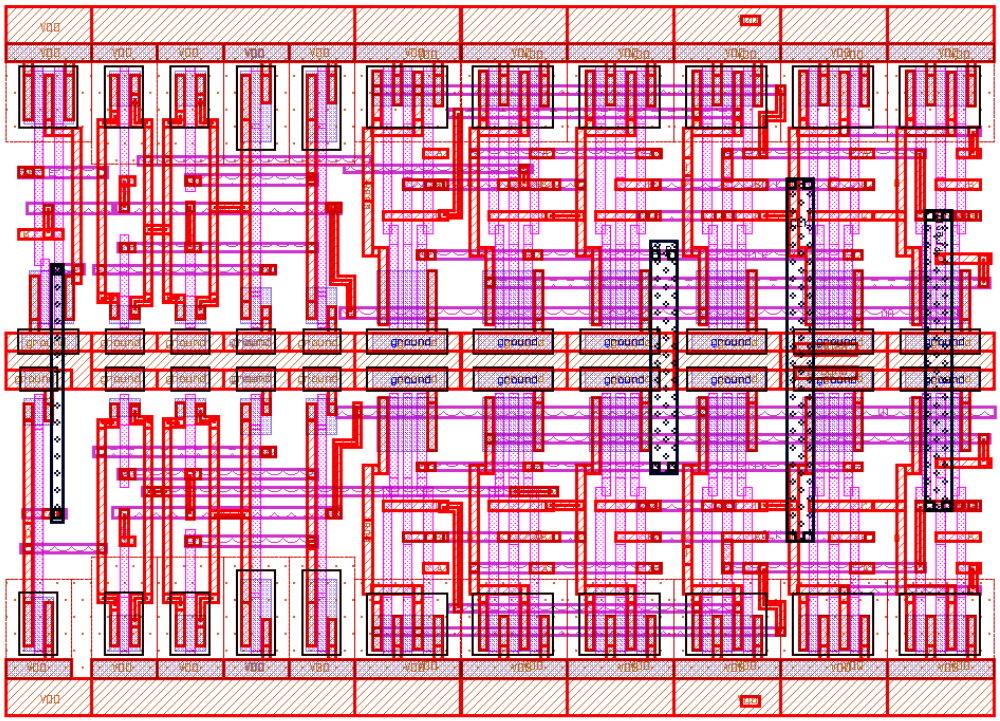


Figure 27: Layout of the 2-bit counter.

6.4. Decoder

The decoder is designed as a one-hot signal to synchronize all four LFSR's. It takes a two-bit binary number from 2-bit counter as input and turns on the corresponding output. The output is connected to the "Reset", "En", "Enable" of LFSRs and decoders, according to Table 3. Input "Disable" is a switch for the whole system. If it is high, the system will be insensitive to any start and stop signal.

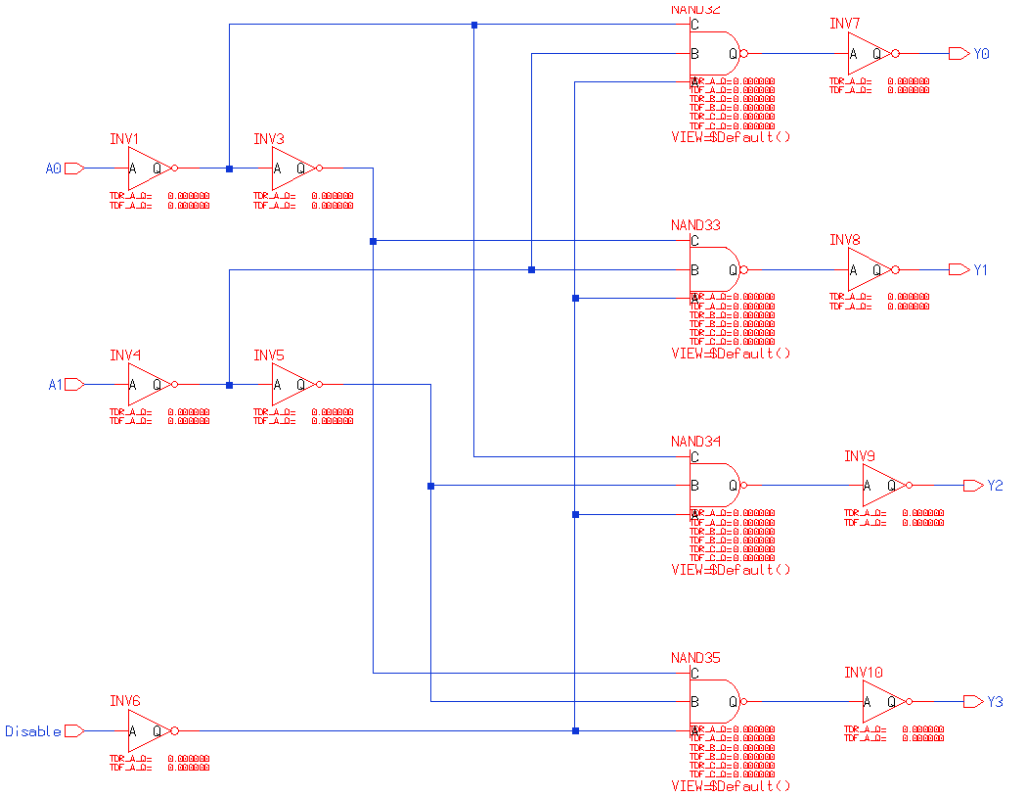


Figure 28: Schematic of the 2-4 Decoder.

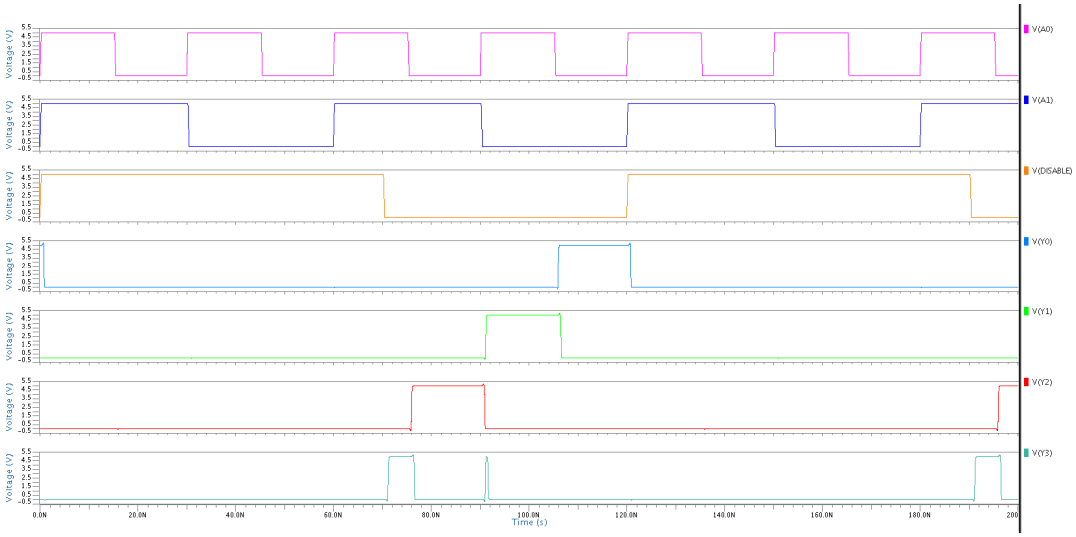


Figure 29: Analog simulation of the 2-4 Decoder.

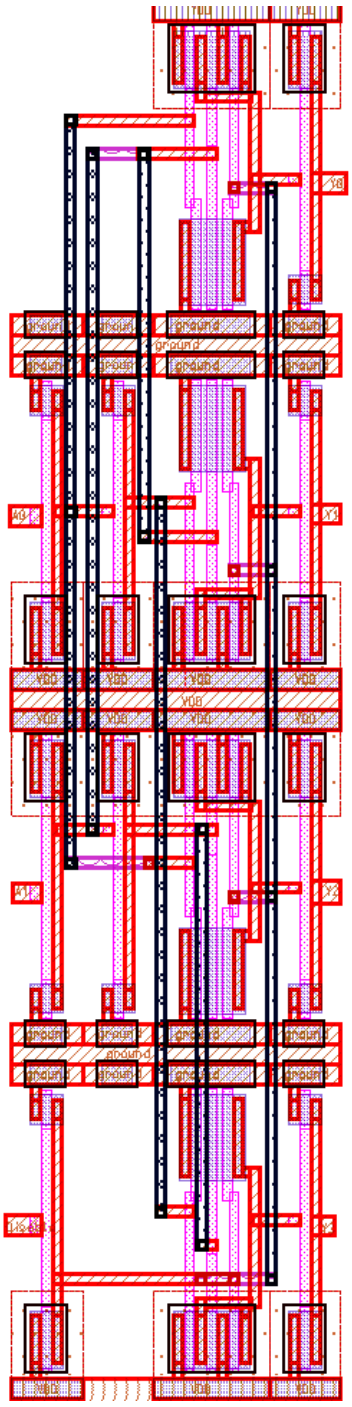


Figure 30: Layout of the 2-4 Decoder.

6.5. LFSR

The main elements of the pulse-detecting circuit are the counting blocks. In the active state, each block counts time from a specified point until a pulse has been detected by the photodiode. We time-multiplex four counting blocks to ensure that the data (time) can be read while the circuit is counting time until the detection of the next pulse. There are two candidates for the counting block: (a) a binary counter and (b) a linear feedback shift register (LFSR). Binary counters have the advantage that their output incrementally increases with each clock cycle, which simplifies the conversion of the counter state to the elapsed time. LFSRs, in contrast, do not produce a sequence of incrementing numbers. Instead, the current state of an LFSR depends on its implementation and on the previous state. Mapping the state of an LFSR to the time requires extra post-processing, e.g. using look-up tables. LFSRs, however, require fewer gates and can operate at higher clock frequencies than binary counters using the same technology. Since our goal is to maximize the time resolution, we use LFSR's to improve performance.

Figure 31 shows the schematic of our 6-bit LFSR implementation, consisting of six D-flip-flops and an XOR gate. This is a modular (or internal XOR) LFSR because the feedback XOR gate is located between the adjacent flip-flops (Bushnell 2000). Modular LFSR's run faster than their counterparts, external-XOR LFSRs, since the path delay

between two adjacent flip-flops is at most that of one XOR gate. An LFSR can be described by its characteristic polynomial:

$$1 + h_1 x + h_2 x^2 + \cdots + h_{n-1} x^{n-1} + x^n,$$

where n is the number of bits and h_i is either 1 if the input of ith flip-flop is XORed with the output of the first flip-flop and 0 otherwise. In our implementation, we use the polynomial:

$$x^6 + x^5 + 1,$$

This is a primitive polynomial, that is, it allows for an LFSR with a maximum length of 2^n-1 = 63 unique states.

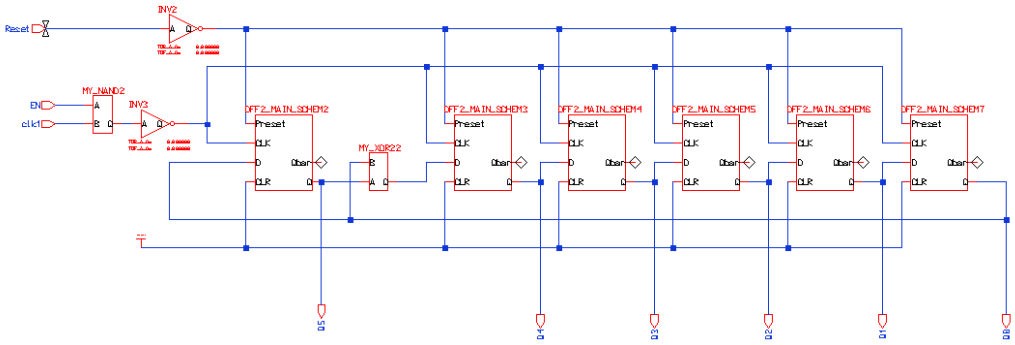


Figure 31: Schematic of the LFSR

Our counting blocks require two additional features: (a) a reset signal (RESET) and (b) an enable signal (EN). If RESET is high, the LFSR is asynchronously set to a

known non-zero state in order to have a reference point for time-interval counting. For this, we use D-flip-flops with a set function to initialize the LFSR to the state "111111". EN is required to stop counting at the time when a photon has been registered and to hold the current value of the LFSR. This function is implemented by gating the LFSR clock with an AND gate.

In order to increase the time resolution of our pulse-detecting circuit, we optimized the LFSR for speed. For this, we compared various implementations of D-flip-flops and decided to implement the LFSR using NAND gates. We varied the widths of the transistors in the NAND gates to optimize for fastest rise- and fall times. Similarly, we optimized the XOR gate to achieve maximum performance.

We performed analog simulations of the LFSR to verify its functionality and determine the maximum clock frequency at which it can operate. Figure 32 shows the voltage waveforms for one of these simulations. When RESET is asserted, the LFSR output 'OUT [5:0]' remains at "111111". When RESET is de-asserted, the LFSR starts generating sequences and 'OUT' changes with every clock cycle. When EN falls down, the LFSR stops counting and holds the value, "001011" in this case. According to the simulation results, the circuit was able to operate at 380 MHz; at higher frequencies, low-to-high transitions failed to propagate through the XOR gate within one clock cycle and the output of the LFSR was stuck at zero. We performed this simulation without a

load at the LFSR outputs; therefore the maximum frequency for the entire system may be lower. Moreover, our circuit model does not include RC-parasitics due to interconnects; therefore the actual maximum operating frequency will be further degraded.

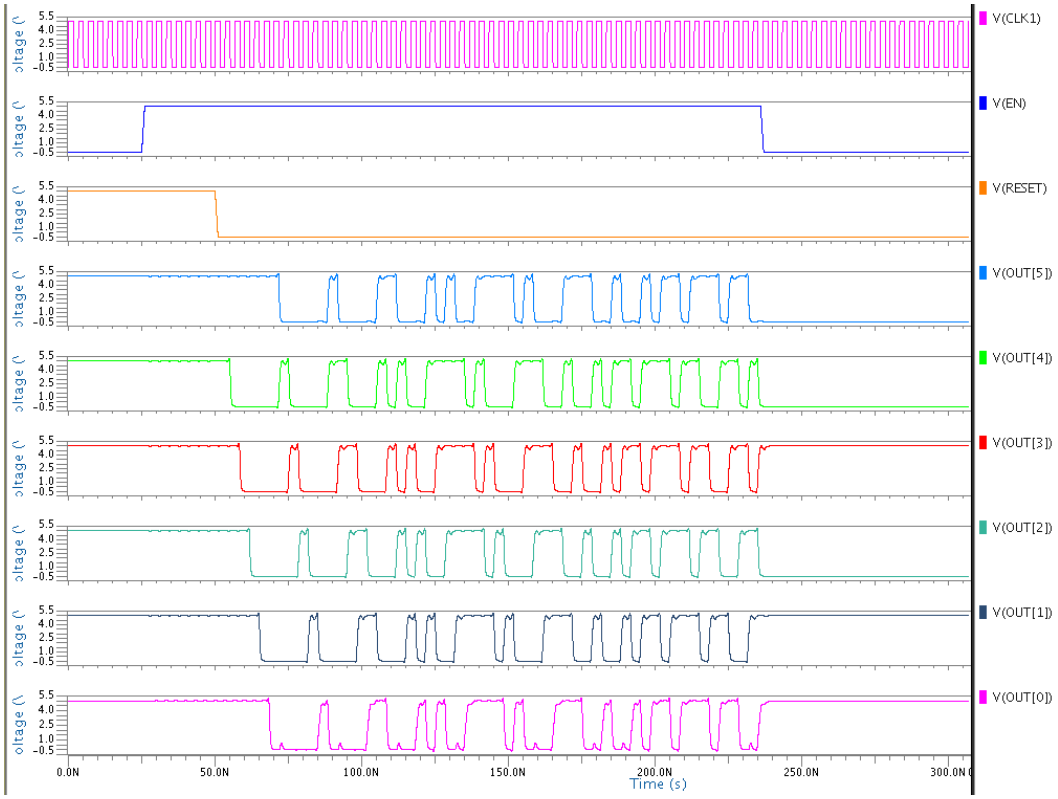


Figure 32: Analog simulation of the LFSR

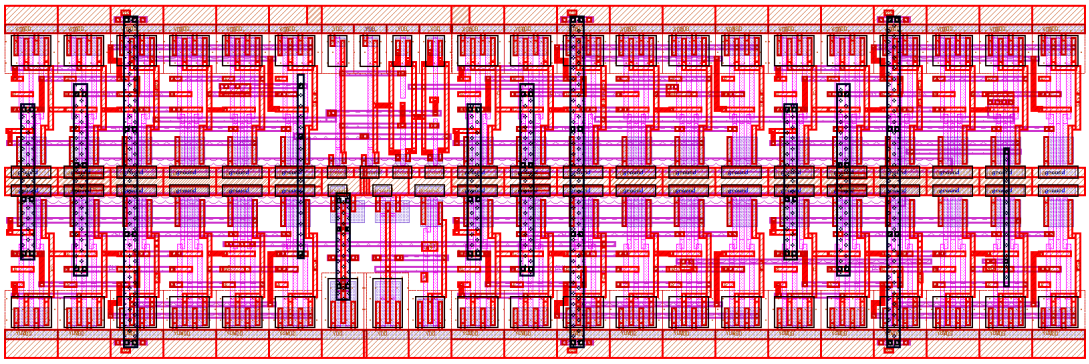


Figure 33: Layout of the LFSR

6.6. MUX

A 4-1 MUX is used in our circuit in order to multiplex the signals coming from the LFSR's. We use a transmission gate implementation for the 4-1 MUX since it uses the least number of transistors compared to any other implementation. For the 4:1 multiplexer, only eight transmission gates and two inverters are needed, a total of 20 transistors. The schematic for the MUX is shown in Figure 34. Here when both the inputs are low, I0 is seen at the output. When the inputs A and B are set to 0 and 1 respectively, I1 is seen at the output and so on. At a given time only a single complete path exists between an input and the output.

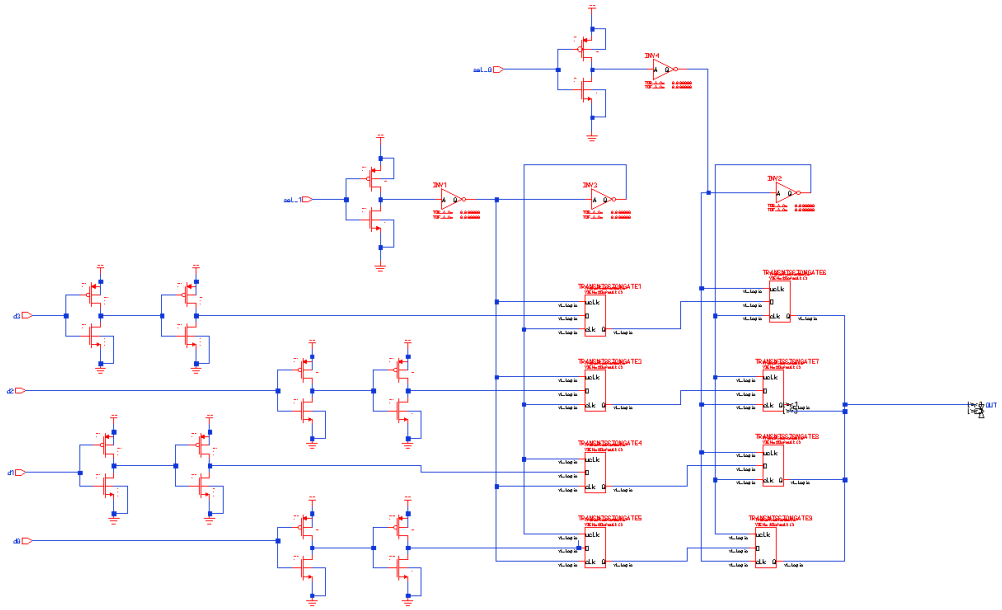


Figure 34: Schematic of the 4-1 MUX

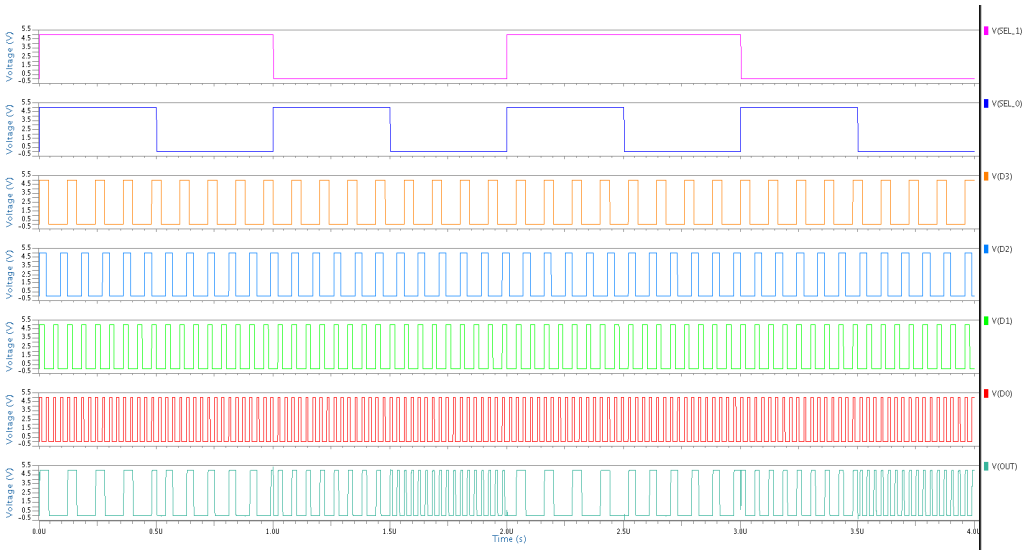


Figure 35: Analog simulation of the 4-1 MUX

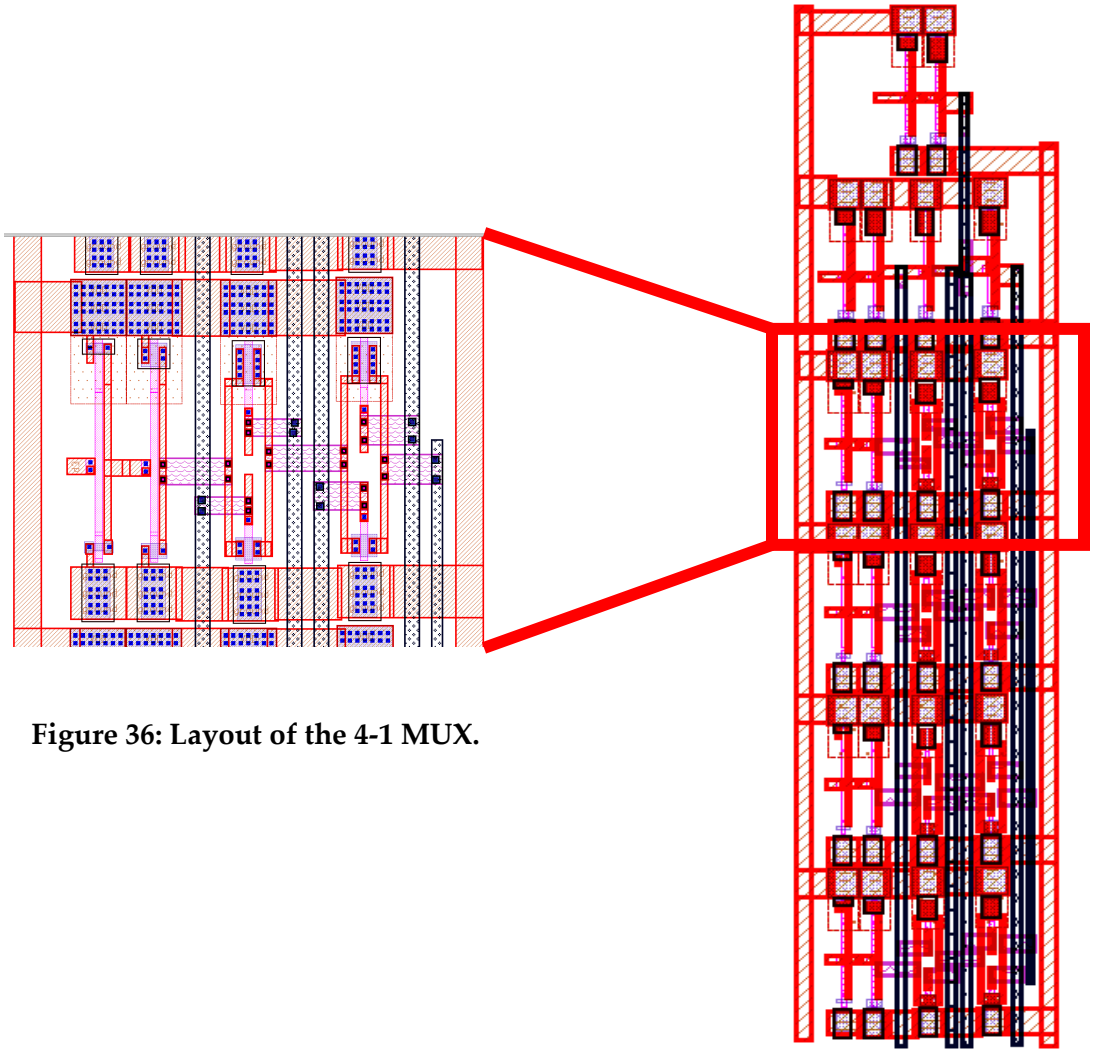


Figure 36: Layout of the 4-1 MUX.

6.7. Results

We performed analog simulations of the entire system to verify its functionality and to determine the maximum operating frequency. Figure 38 shows the waveform of the top-level simulation. CLK2 is an extra clock, which is 30x slower than CLK1, is used to schedule the time frames of the LFSRs. SIGNAL represents the incoming pulse which

sets the FLAG in the next time frame. The LFSR holds the value “000101”; this state occurs 22 clock cycles after the initial state and we can trace back the time when the pulse was registered by the photodiode.

We were able to run top-level simulation at the speed of 370Mhz. This is slightly lower than the frequency of the LFSR due to extra load on the LFSR.

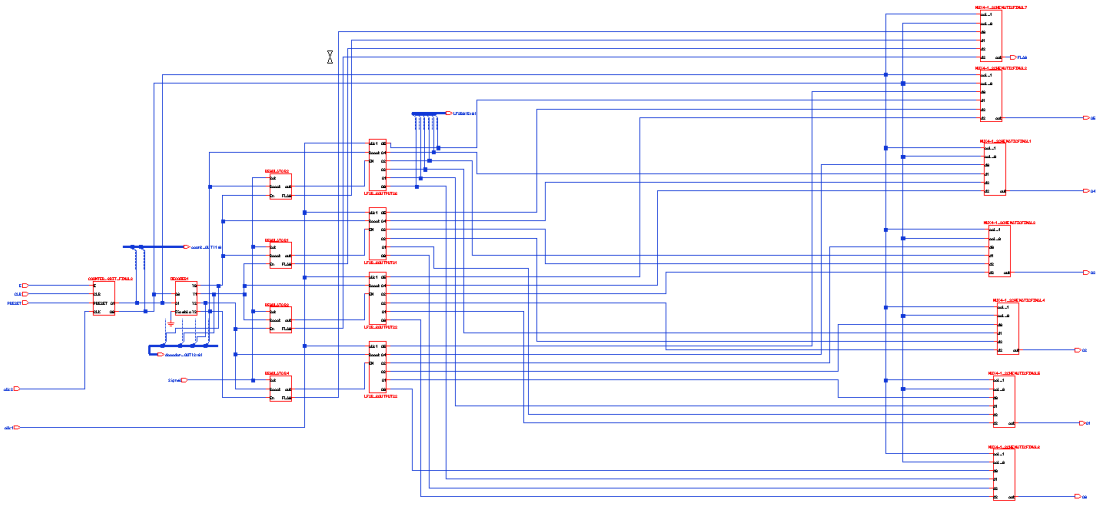


Figure 37: Top level schematic

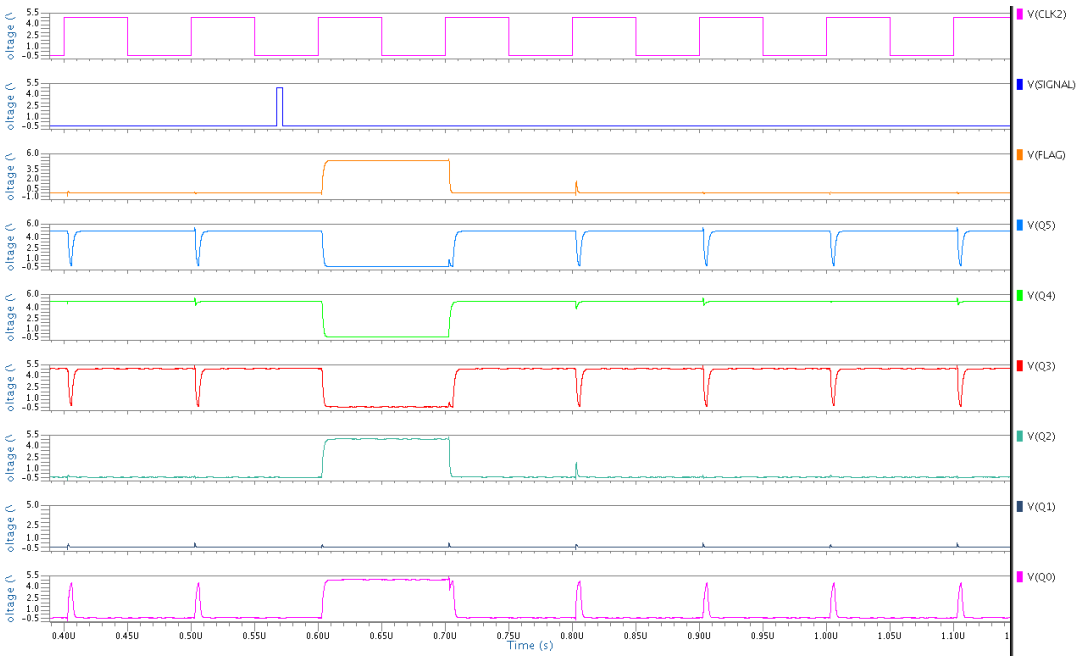


Figure 38: Top level analog simulation

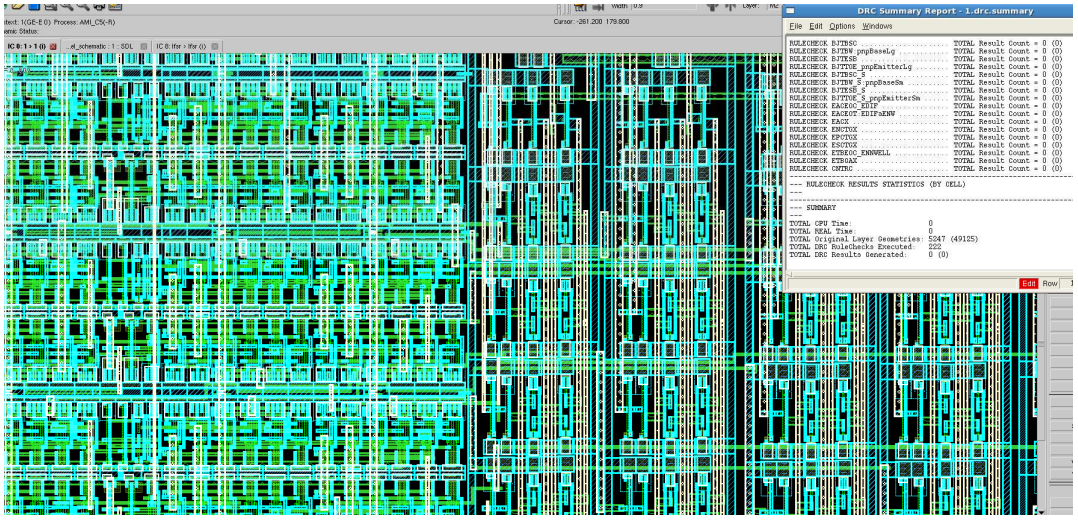


Figure 39: DRC for top-level layout

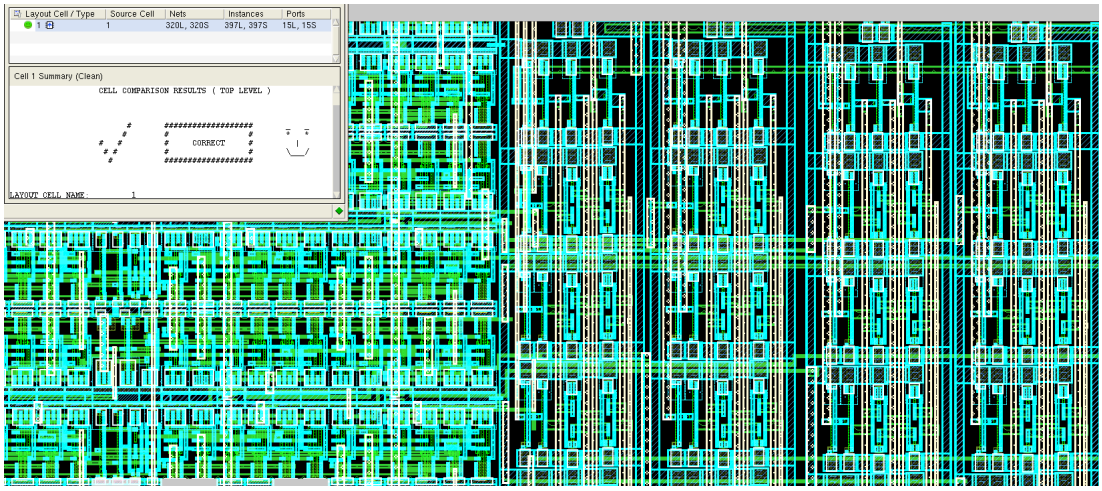


Figure 40: LVS check for top-level layout

6.8. Area Estimate

Our design uses a total of 1492 transistors and has a total area of 0.495 sq. mm. The breakdown of this estimate is as follows: The D flip-flop uses 6 3-input NAND gates resulting in a total of 36 transistors and an area of 2540 sq. microns. The counter uses 2 D flip-flops, 2 2-input XOR gates and 1 2-input AND gate resulting in 94 transistors. Each regulator uses 2 inverters, 2 Pseudo-NAND gates and 2 NAND2 gates resulting in a total of 80 transistors. The Decoder uses 4 3-input NAND gates and 9 inverters resulting in a total of 42 transistors. Each LFSR uses 6 D flip-flops, 3 XOR gates, 2 inverters and 1 NAND gate resulting in a total of 968 transistors. Finally, each MUX uses 8 transmission gates and 28 inverters resulting in a total of 308 transistors.

6.9. Power Consumption

The dynamic power consumption of our chip with a 370 MHz clock is given by the following:

$$P = \alpha C V_{DD}^2 f = 0.1 * [1492 * (12\lambda) * \left(\frac{0.8\mu m}{2\lambda}\right) * 2fF/\mu m)] * 5^2 * f$$

= 13.2 mW

We therefore designed and simulated the readout and processing architecture for single photon avalanche photodiode detectors. To the best of our knowledge, this is the first demonstration of a design using digital elements for processing the output of a single photon avalanche photo diode. Our design in 0.5 μm technology operates at 370 MHz clock frequency while enabling parallel readout. Using this architecture, we can process every detected signal, which enables four times shorter integration times. The architecture is scalable and therefore amenable to be used in large integrated arrays of photo detectors. Our novel design can be used where good timing accuracies, parallelism, and small areas are required.

7. Generation and Optimization of Signatures

In order to generate and optimize the signatures of our RET keys, we make use of 3 randomly chosen networks as shown in **Figure 41**. The networks are excited using 488 nm, 543.5 nm and 647 nm at various time delays. The fluorescence from all three fluorophores was recorded and compared. The values of R_0 between the fluorophore pairs are shown in **Table 5**.

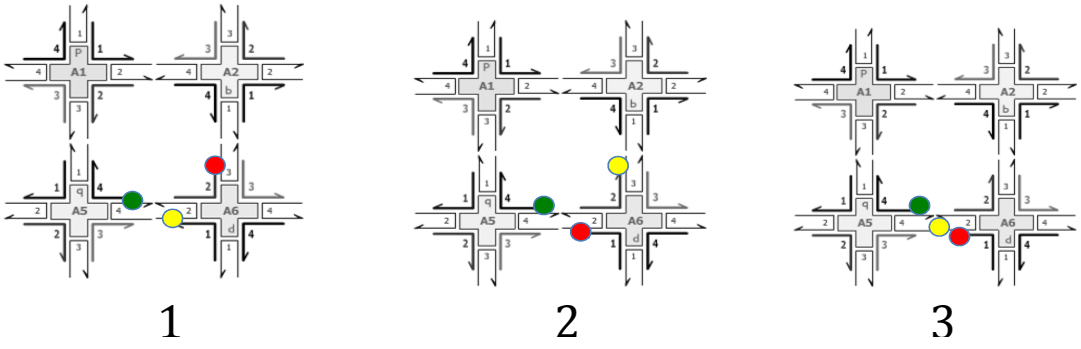
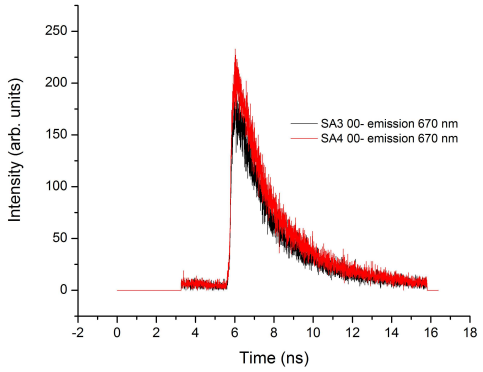


Figure 41: RET networks with fluorophores AF 488 (green), AF 594 (yellow) and AF 647 (red). In these networks, AF 488 and AF 594 serve as the donors and AF 647 serves as the acceptor. The proximity of the acceptor from its nearest donor decreases from 1 to 2 to 3. Figure not to scale.

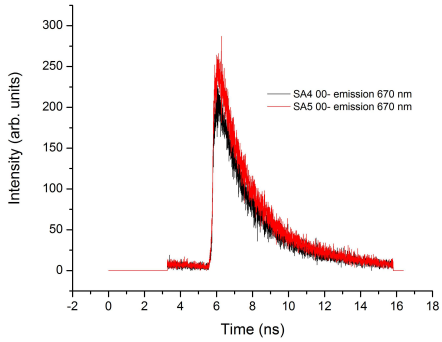
Table 5: The R_0 values between the fluorophore pairs used in our RET networks.

	AF 488	AF 594	AF 647
AF 488	4.87E-09	5.89E-09	5.42E-09
AF 594	1.83E-09	5.58E-09	7.80E-09
AF 647	7.37E-10	2.29E-09	6.62E-09

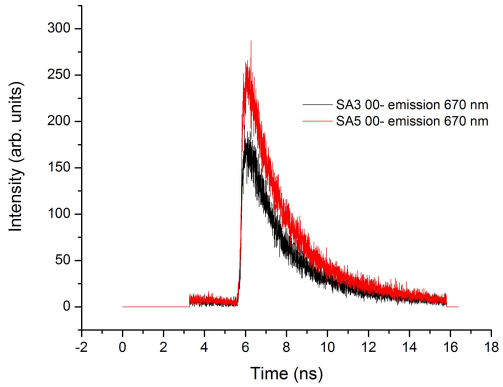
The raw TCSPC data for all 3 networks is shown in Figure 42. For the generation of these histograms, the networks were excited simultaneously at 488 nm and 543.5 nm and observed at 670 nm. Therefore, AF 488 and AF 594 served as donors and AF 647 served as the acceptor in these networks. We can see that the proximity of the donors to AF 647 decreases as we go from sample 1 to 3. In sample 1, AF 647 is nearly 10 nm away from the donors and as a result the output fluorescence is low. We notice an increase in the amplitude of the signal as we go from sample 1 to 2 as expected, and a further increase in sample 3 where AF 594, the donor, is 4.49 nm away from AF 647. We therefore see that the TCSPC results of these networks confirm with the predicted trends from the Förster's equation as described in chapter 3. We now look at various ways to quantitate the difference in signatures from these networks.



(a)



(b)



(c)

Figure 42 (a): As expected, the fluorescence from AF 647 in sample 2 is higher than that from sample 1. This is due to the higher proximity of AF 647 to its donors in sample 2. (b). Fluorescence from AF 647 in sample 3 is higher than that of sample 2 since AF 594 is much closer to AF 647 in sample 3 and therefore transfers more energy. (c) In this figure, fluorescence from AF 647 in sample 3 is higher than sample 1, as expected.

7.1 Pearson Correlation Coefficient

In order to compare the signatures of the 3 networks shown in Figure 41, we considered using the method of least squares, the Pearson correlation coefficient, the Spearman correlation coefficient and the Kendall correlation coefficient. We found that Pearson correlation coefficient gave the highest sensitivity to sample signatures and use of this method resulted in maximum difference between the intra-key and inter-key correlations.

The Pearson correlation coefficient is the most commonly used metric for comparing two signatures when a linear relationship exists between them (Snedecor and Cochran 1989). Perfectly correlated signatures (increasing linear relationship) are assigned a correlation coefficient of 1 while -1 is assigned to signatures that are perfectly anti-correlated (decreasing linear relationship). A correlation coefficient of 0 indicates that the two signatures are uncorrelated. We can use the Pearson correlation coefficient since the signatures of our samples satisfy the following conditions (Snedecor and Cochran 1989):

1. The two samples being compared are measured independently.
2. The values being compared are measured, not controlled
3. The covariation between the two samples is linear.

4. The two signatures being compared are exponentially distributed. Pearson correlation coefficient requires that the two random variables being compared have a Gaussian distribution. Exponential distribution can be approximated as a Gaussian distribution with excess kurtosis (Chok 2010).

We calculate the Pearson correlation coefficient (ρ) using the following formula:

$$\rho_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sqrt{E[(X - E(X))^2]}\sqrt{E[(Y - E(Y))^2]}}$$

We applied the above formula to the semi-log plot of the three signatures shown in Figure 42. A semi-log plot of the raw histogram minimizes the effect of the excitation pulse on the correlations and enables the comparison of the output fluorescence from the three samples. We obtained 100% correlation for similar keys under similar excitation conditions. However, when different keys were compared under similar excitation conditions, we again noticed high correlations with the mean of the Gaussian fit at 96.71% and a FWHM of 52.71%. This result, shown in Figure 43, was expected given the poor discrimination between the signatures for the different networks. There are 2 primary reasons for the poor discrimination:

1. We do not filter out the noise components in the signal and the excitation pulse from the raw histogram. This results in samples with different exponential decays correlating very highly.

2. The signal to noise ratio of the histograms is too low. This results in the amplitude values of the various networks falling within a narrow range resulting in high correlations.

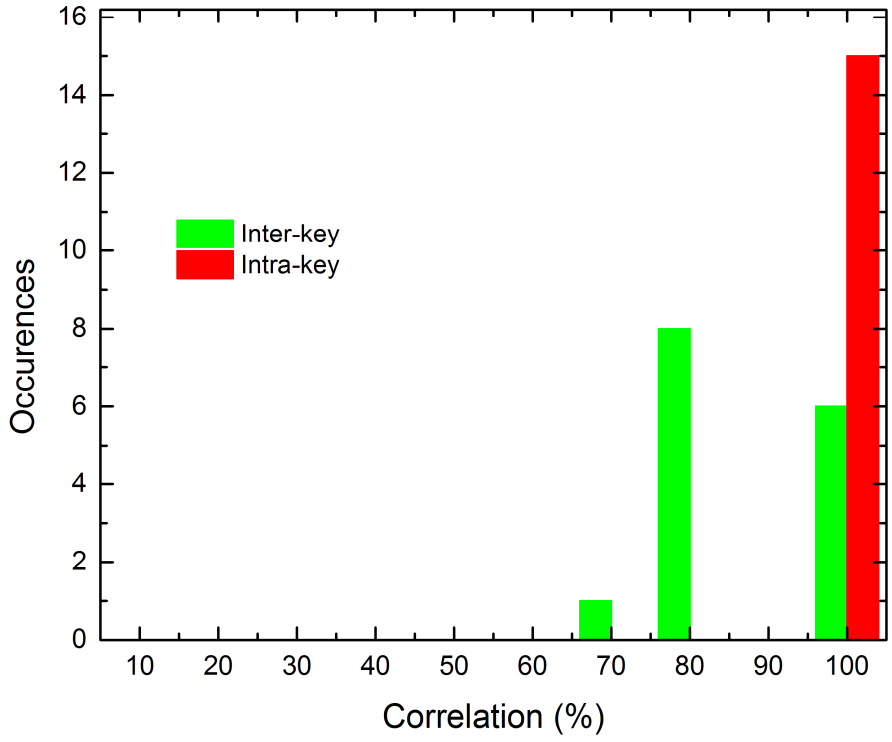


Figure 43: Intra-key and Inter-key correlations of signatures of networks 1, 2 and 3. We see that the intra-key as well as inter-key correlations are high.

7.2. Hough Transform

In order to optimize the signatures for increased sensitivity, we compare features in the histogram that are unique to the sample such as lifetime and amplitude information while removing the local fluctuations present in the signal as well as the contribution from the excitation pulse.

Traditional techniques to measure the lifetime of a sample involve measuring the dominant slope in the fluorescence decay. However, detecting a single slope accurately is often difficult due to the multi-exponential nature of the decay curve, local fluctuations in the fluorescence decay that result from the randomness associated with fluorescence and the contribution from the excitation pulse to the output. We refer to the contribution from these three sources as 'noise' in the signal. We therefore make use of a feature extraction technique, known as the Hough transform that can filter out the noise and extract the dominant lifetimes present in the sample. To use this technique, we first convert the multi-exponential fluorescence decay to a semi-logarithm image and then extract straight lines corresponding to the lifetimes present in the output fluorescence signal. The resultant signal is unique to the combination of the excitation conditions and the key being probed. A more detailed description regarding the implementation of this technique follows.

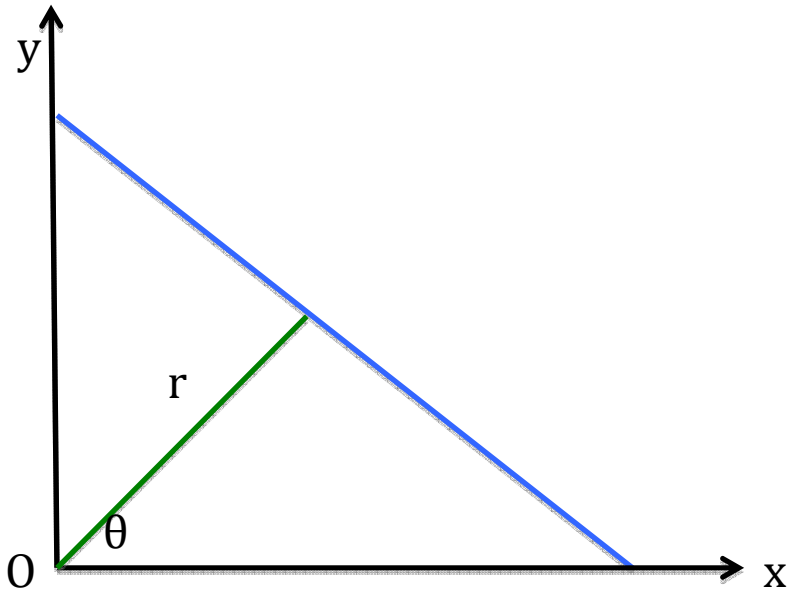


Figure 44: The Hough transform calculates *r* and θ for the family of lines generated through every point on the histogram (Duda and Hart 1972).

The Hough algorithm generates a family of lines at different angles through each point of the semi-logarithm histogram. Each of these lines can be expressed using the angle and distance from a fixed point on the histogram. In the above image, the parameters for the blue line are established by drawing the shortest perpendicular from the line to an origin as indicated by the green line. The length of the perpendicular is taken as '*r*' and the angle of this perpendicular with respect to the horizontal is taken as 'θ'. The equation for the blue line is written as $r = x\cos\theta + y\sin\theta$ in the polar coordinate system. For each line generated by the Hough transform for any given data point, the shortest perpendicular from the line to the origin is drawn and values of *r* and θ calculated. This procedure is carried out for all the points in the histogram. A two-

dimensional array, known as an accumulator, with quantized angles along one dimension and a quantized distance along the other is incremented with the (r,θ) value of every line drawn through every point on the histogram. The (r,θ) value with the maximum number of votes is established as the Hough peak and a Hough line is generated by connecting all the data points with a common (r,θ) value. The tangent of (90 - θ) value of the Hough peak determines the lifetime of the sample being examined. The use of the Hough transform helps us eliminate the noise contribution to the signal since noise is unlikely to dominate any one bin.

The code for signature generation and optimization may be found in Appendices B, C. The details regarding the optimization parameters are briefly explained here.

We first normalize the two histograms being compared in order to correct for inadvertent differences in the signatures of identical keys. Despite the use of the identical fabrication protocols, measurement techniques and conditions, some differences might result in the measurement of different batches of the same RET key. This happens due to several non-idealities that are introduced during the fabrication and characterization of the RET keys. Varying the raw material and structural yields of the DNA grids could introduce variations during the fabrication process while variations in the temperature of the filter, cuvette replacement error and minor pipetting variations could introduce errors during measurements of the keys.

After normalization of counts, each semi-logarithm histogram is scaled by a factor of 200 in order to compensate for the compression that results from converting the data to a logarithmic scale. Subsequently, this histogram is converted into a black and white image with the non-zero, positive values being assigned a 1 and all other values being assigned a 0. We then use MATLAB's built-in function, 'hough' with 'RhoResolution' 5 and 'Theta' -90 to 89.8 in steps of 0.2, in order to calculate the Hough transform of the image. The 'peaks' and 'lines' functions in MATLAB are then used to calculate the dominant Hough peak and the corresponding Hough line with 'Minlength' set to 200 in order to eliminate the noise peaks. The slope of this Hough line will result in the lifetime of the key. We use this lifetime information to compute the regions to be compared in the Hough matrices of the two samples. We calculate the angles in the Hough matrix to be compared using the following formula: Angle = (round (90-arctan (Slope))*5. We found that in order to account for noise, comparing angles $\pm 11°$ resulted in higher intra-key correlations and more accurate inter-key correlations.

In order to verify the sensitivity of the Hough transform as a feature extraction technique, we measured the change in correlation for small changes in lifetime and amplitude. In the first experiment, we generated two exponentially decaying curves of equal amplitude but with 0.1 ns difference in lifetime as shown in Figure 45. We then applied the Hough transform to extract the lifetimes of each of the curves and then

calculated the Pearson correlation coefficient for the two curves. The two curves differed by 32.72% even for a 0.1 ns change in the lifetime.

In the second experiment, we generated two exponentially decaying curves with the same lifetime but with a 10% change in amplitude. As shown in Figure 46, we noticed a 36.42% change in the signatures between the two samples when we applied the Hough transform and calculated the Pearson correlation coefficient between them.

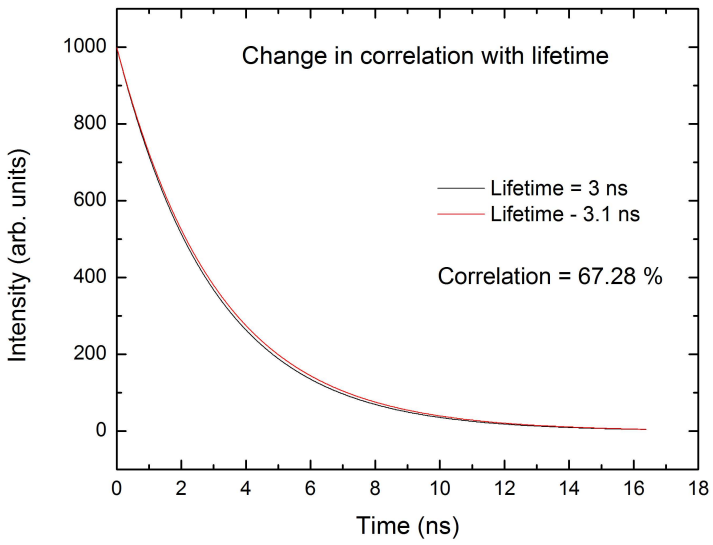


Figure 45: The figure shows the sensitivity of the Hough transform to minor variations in lifetime. We notice a 32.72% change in correlation on changing the lifetime by 0.1 ns while keeping the amplitude the same.

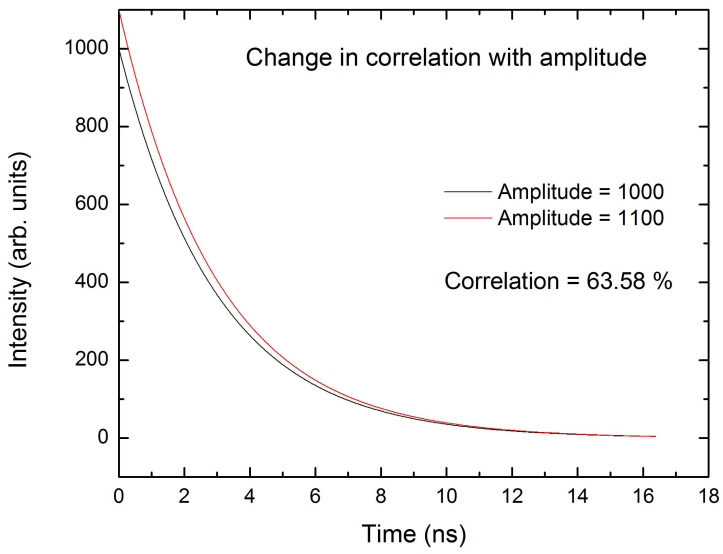


Figure 46: The figure shows the sensitivity of the Hough transform to minor variations in amplitude. We notice a 36.42% change in correlation on changing the lifetime by 10% while keeping the lifetime of both the decay curves the same.

Since the Hough transform method proved to be very sensitive to minor variations in the signatures, we applied it to the signatures shown in Figure 42. The Pearson correlation coefficients for the intra-key and inter-key comparisons are shown below. The Hough angles 11 to 63 degrees correspond to lifetimes of 0.5 ns to 6 ns, which is the range of lifetimes we see in the samples 1, 2 and 3. Compared to the semi-logarithm signatures, the intra-key correlations now have a mean of 65.19% and a FWHM of 5.17%. The inter-key correlations substantially improve and have a reduced mean of 34.1% and a FWHM of 5.71%. We clearly notice an increase in the separation between the intra-key and inter-key correlations when compared to Figure 43.

The intra-key correlations between the signatures are lower than the ideal values. We hypothesized that this was the result of low signal to noise ratio on the histograms. In order to optimize our signatures further, we made use of the lens arrangement, shown in Figure 17, to focus the output fluorescence from our sample onto the SPAD. As a result, we noticed 3 orders increase in the magnitude of the signal that could be detected on TREX. The graphs below show that we can easily discriminate between the signatures of dissimilar keys. The increase in the signal to noise ratio resulted in significantly higher intra-key correlations, which further increased the separation between the intra-key and inter-key correlations.

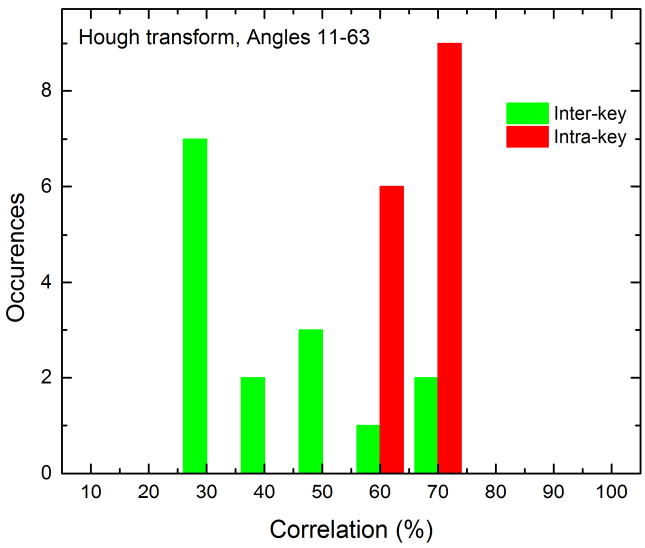


Figure 47: Intra-key and inter-key correlations of the signatures from the networks shown in Figure 43 on applying the Hough transform with select angles. A further separation between the intra-key and inter-key correlations is noticed but the intra-key correlations continue to be low.

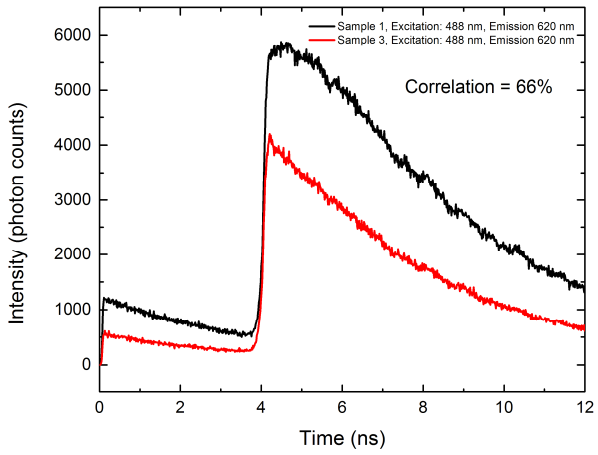


Figure 48: The figure shows the signatures of networks 1 and 3 under identical excitation conditions. Optimizing the collection efficiency of TREX enabled increased discrimination between signatures.

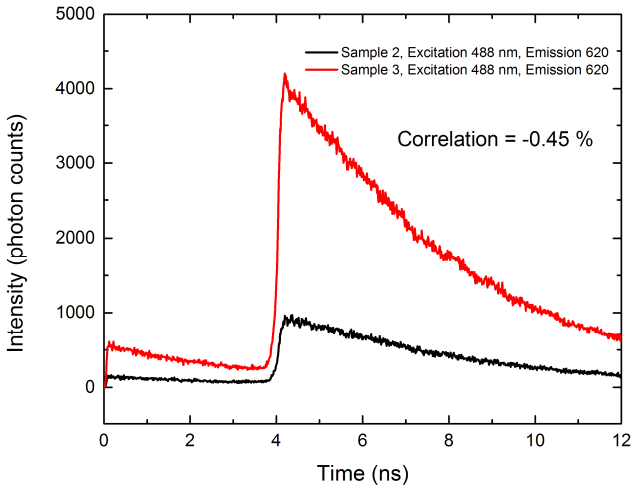


Figure 49: The figure shows the signatures of networks 2 and 3 under identical excitation conditions. Optimization the collection efficiency of TREX enabled the increased discrimination between signatures.

We compared the separation between the intra-key and inter-key correlations, or the Hamming distance, of the various signature-generating schemes. Figure 50 shows that using the semi-log histogram as the signature provided the least difference between the intra-key and inter-key correlations with an average Hamming distance of 16%. Using the Hough transform with angles 11-63 degrees combined with the optimized set-up allowed for the best discrimination between the intra-key and inter-key correlations with and average Hamming distance of 39.5%.

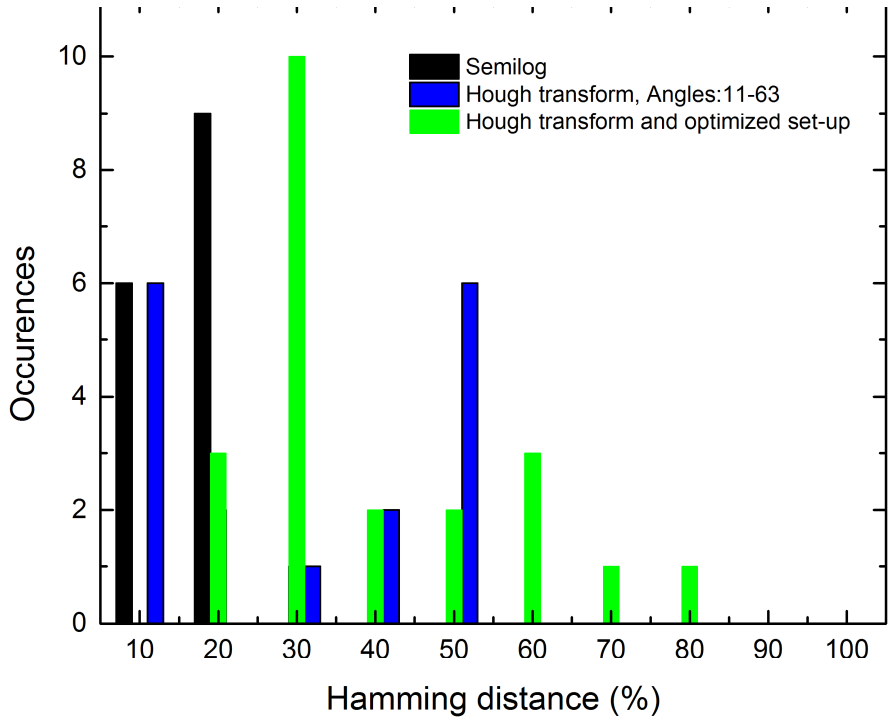


Figure 50: Hamming distance is the separation between the intra-key and inter-key correlations. The graph shows that using the semi-log histogram as the signature provided the least difference between the intra-key and inter-key correlations while use of the Hough transform with angles 11-63 degrees combined with the optimized set-up allowed for the best discrimination between the intra-key and inter-key correlations.

We also noticed that the correlations varied considerably based on the observation wavelength. Figure 51 shows the variation in the output correlations when fluorescence was observed from AF 594 using the 620 nm filter and from AF 647 using the 670 nm filter. Clearly, the correlations using the 620 nm filter are significantly different in terms of intra-key and inter-key correlations. It must be emphasized that the

network being probed and the excitation conditions are exactly the same in both cases and that only the observation wavelength is changed.

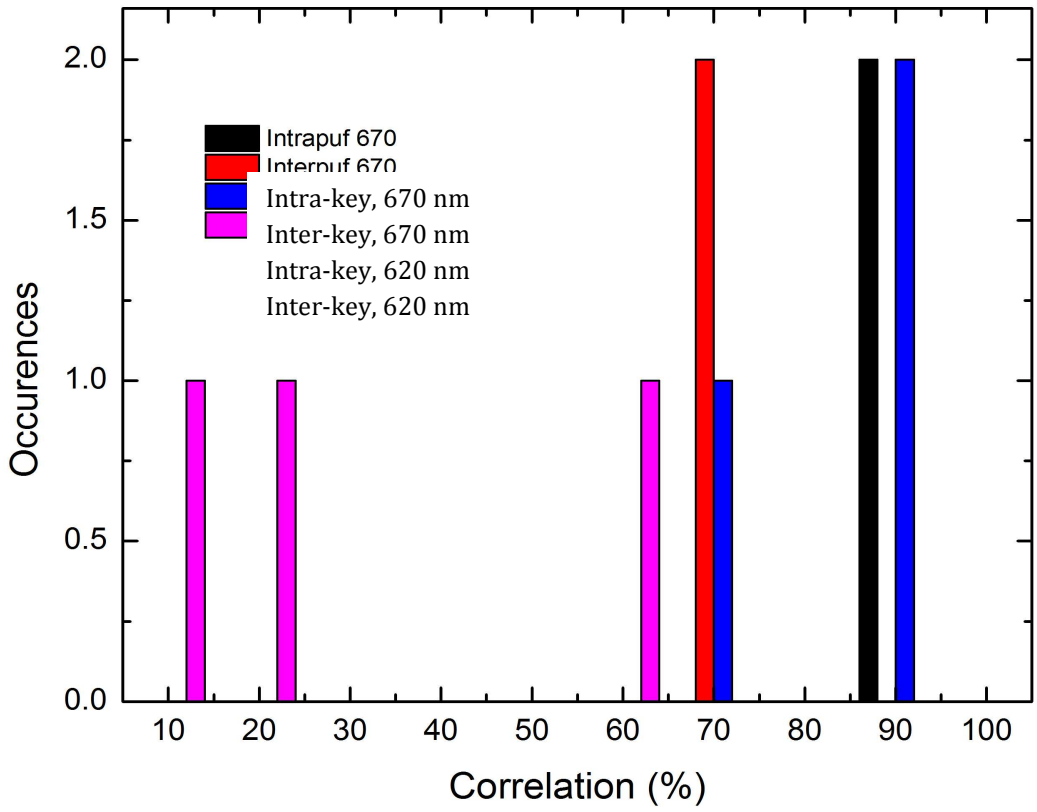


Figure 51: Variation of correlations for the same key with change in observation wavelength. The figure demonstrates that the response of the same key can be varied significantly by changing the challenge.

This result demonstrates an important property of the key: varying a part of the challenge can significantly change the response of a key. This is important because of three reasons:

1. We do not want a key to result in the same signature every time it is probed, since an adversary could use this signature to correlate the behavior of the user across multiple devices. Additionally, a single signature would be easy to steal and immediately allows the adversary access to all the devices the user uses.
2. The total number of responses from our key would now be a combination of the total number of possible challenges and keys.
3. If an adversary steals a key and the CRP's are exchanged through a private channel, he would still have to go through all possible challenges, in the correct order, to find the right response. However, if the challenges are public and the obscurer is computed before every authentication attempt, the adversary would be able to decipher the correct response.

While correlation between signatures helped us identify the optimal parameters for our set-up and the data analysis script, it is not an ideal metric for authentication. When correlation is used, the variation in intra-key correlations changes as ≈±10%. This implies that an adversary only has to try 9 values before he is successfully authenticated 100% of the time. We could technically limit the number of authentication attempts that a user is allowed but the probability of a random guess resulting in successful authentication is as high as 11.11%. In order to significantly lower the high probability of false positives, we use the Hamming distance between signatures for authentication. For this purpose, we use the procedure described above but instead of calculating the Pearson correlation

coefficient, the fluorescence histograms are converted to a bit string and then compared. In the bit string, if the number of positions in which the bits differ is below a certain pre-determined threshold, authentication is established. We found that converting the Hough transforms of the fluorescence histogram into a bit string using a threshold of nine achieved the best results when the distance between the intra and inter-key Hamming distance was used as the metric. We also found that converting the raw histogram into a bit string with positive values corresponding to a 1 and the rest to 0's, did not give us meaningful results when we observed the trends in the intra and inter-key correlations. We attribute this to the shift in the bin where the non-zero value is located with otherwise negligible variations in amplitude or lifetime. Similarly, converting the Hough matrix to a bit string with positive values corresponding to a 1 and the others to 0 did not result in high correlation for similar histograms and low correlations for dissimilar histograms. This was expected since we did not eliminate the noise bins in the Hough matrix. The Hough matrix consists of some peaks corresponding to the dominant lifetimes and amplitudes in the sample and others that are generated largely due to noise. We noticed that the bins corresponding to the peaks representative of the sample have values in the high 20's and low 30's while peaks that resulted from noise generally have values less than 9. We therefore assigned all bins with values less than 9 to 0 while the bins with values greater than 9, were assigned a 1.

This thresholding resulted in high correlations for similar histograms while dissimilar histograms correlated poorly.

In summary, we established a method to evaluate and compare the signatures of the RET-key. The Hough transform is extremely sensitive to small changes in the signatures of RET-keys. It efficiently filters out noise from the output signal and enables the comparison of pre-specified lifetimes between two fluorescence signals. Converting the transformed data into a bit string resulted in a low Hamming distance between similar keys under similar excitation conditions and a high Hamming distance between dissimilar keys under dissimilar excitation conditions. In the following chapter, we will use the technique described in this chapter on a larger set of keys and excitation conditions to establish whether we still see a low Hamming distance between similar keys under similar excitation conditions and a high Hamming distance between dissimilar keys under dissimilar excitation conditions.

8. Survey of RET-Keys

We carried out a detailed survey of three-fluorophore RET-keys under a large number of excitation and emission conditions and analyzed the intra-key and inter-key behavior of the responses. The purpose of this study was four fold: (1) To conclusively demonstrate, from experimental results, that the intra-key correlations are high and well separated from the inter-key correlations. Also, to verify whether the intra-key correlations have a narrow distribution while the inter-key correlations are distributed over a wide range, (2) To identify the number of collisions in the output space using the simplest possible network of fluorophores, since increasing the complexity of the networks will only increase the total number of outputs. The total number of collisions is an important metric to establish the strength of the key against various cryptographic attacks as will be discussed in chapter 10, (3) To study the time-evolution of two identical keys manufactured separately under identical excitation conditions, and (4) To measure the longevity of the RET-key under normal measurement conditions, i.e., no effort was made to modify the rate of change of the signature either by modifying the concentration of the sample or intentionally bleaching it.

In order to carry out the study, we performed four separate experiments, which are listed below:

1. In the first experiment, the network was changed but the excitation conditions remained unchanged.
2. In the second experiment, the excitation wavelength was varied but the excitation delay and network were unchanged.
3. In the third experiment, the excitation delay was varied but the excitation wavelength and network were unchanged.
4. In the last experiment, the time-response of two batches of the same key was observed under identical excitation conditions.

This chapter describes the details regarding these experiments, the experimental results and conclusions.

8.1. DNA sequence design

The nanoscale keys are constructed using hierarchical DNA self-assembly as described in section 2.1. Each DNA grid used in the surveys consists of 4 tiles and is 40 nm x 40 nm in size. Synthetic oligonucleotides purified using HPLC were purchased from Integrated DNA Technologies (Coralville, IA). The sequences of the native and fluorophore-functionalized strands are given in Table 6.

Table 6: Sequence and conjugation information of DNA strands used in the experiments.

Strand Name	Sequence
CoreA	aggcaccatcgtaggttttcgttgcgatcaccaacggagttttttctgccgtacaccagt gaagtttttcgatcctagcacctctggagtttttcttgcc
SA1	atgcaacctgcctggcaagactccagaggactactcatccgt
SA2	tccgactgagccctgctaggatcgacttcactggaccgttctaccga
SA3	accggaggcttcctgtacggcagaactccgttggacgaacag
SA4	atagcgcctgatcgcaacgcctacgatggacacgccg
Arm 1.1	gttatcggcgtgtggttgcataatac
Arm 1.2	caatcacggatgagtagtgggctcagtcggacattc
Arm 1.3	cctcgtcggtagaacggtggaagcctccggtcgtgc
Arm 1.4	ttcaactgttcgtggcgctatattgt
Arm 2.1	caagccggcgtgtggttgcatacgac
Arm 2.2	aagtgacggatgagtagtgggctcagtcggatactg
Arm 2.3	ttgattcggtagaacggtggaagcctccggtttaca
Arm 2.4	gattgctgttcgtggcgctatgaatg
Arm 5.1	cgaggcggcgtgtggttgcatgcacg
Arm 5.2	ttaagacggatgagtagtgggctcagtcggattgta
Arm 5.3	tcatgtcggtagaacggtggaagcctccggtttgct

Arm 5.4	tgtagctgttcgtggcgctattacgt
Arm 6.1	tctgacggcgtgtggttgcattcaac
Arm 6.2	ctacaacggatgagtagtgggctcagtcggaacgta
Arm 6.3	gcttgtcggtagaacggtggaagcctccggtgtcgt
Arm 6.4	taacgctgttcgtggcgctatcattg
Arm 6.2 (AF 488)	/5alex488n/ctacaacggatgagtagtgggctcagtcggaacgta
Arm 6.2 (AF 594)	/5alex594n/ctacaacggatgagtagtgggctcagtcggaacgta
Arm 6.2 (AF 647)	/5alex647n/ctacaacggatgagtagtgggctcagtcggaacgta
SA1 (AF 488)	atgcaacctgcctggcaagactccagaggactactcatccgt/3alexf488n/
SA1 (AF 594)	atgcaacctgcctggcaagactccagaggactactcatccgt/3alexf594n/
SA1 (AF 647)	atgcaacctgcctggcaagactccagaggactactcatccgt/3alexf647n/
SA2 (AF 488)	tccgactgagccctgctaggatcgacttcactggaccgttctaccga/3alexf488n/
SA2 (AF 594)	tccgactgagccctgctaggatcgacttcactggaccgttctaccga/3alexf594n/
SA2 (AF 647)	tccgactgagccctgctaggatcgacttcactggaccgttctaccga/3alexf647n/
SA4 (AF 488)	/5alex488n/atagcgcctgatcgcaacgcctacgatggacacgccg
SA4 (AF 594)	/5alex594n/atagcgcctgatcgcaacgcctacgatggacacgccg
SA4 (AF 647)	/5alex647n/atagcgcctgatcgcaacgcctacgatggacacgccg

8.2. Capillary Gel Electrophoresis Results

DNA strands pre-conjugated with fluorophores are purchased from IDT-DNA. The conjugation yield of the fluorophores to the DNA strands is verified using Capillary Gel Electrophoresis (CGE). A polyacrylamide capillary (10% monomer) is used with an injection time of 10 seconds at -10 KV and a run time of 65 minutes at -12.7 KV. Absorbance is measured at 258 ± 6nm. We first run CGE on a DNA ladder that consists of six DNA strands with the following lengths: 25, 26, 27, 28, 29 and 30 bases. As shown in Figure S1, we successfully resolved six peaks corresponding to the six different lengths of single-stranded DNA in the ladder sample. This indicates that we obtain single base resolution using our CGE set-up and the parameters mentioned above.

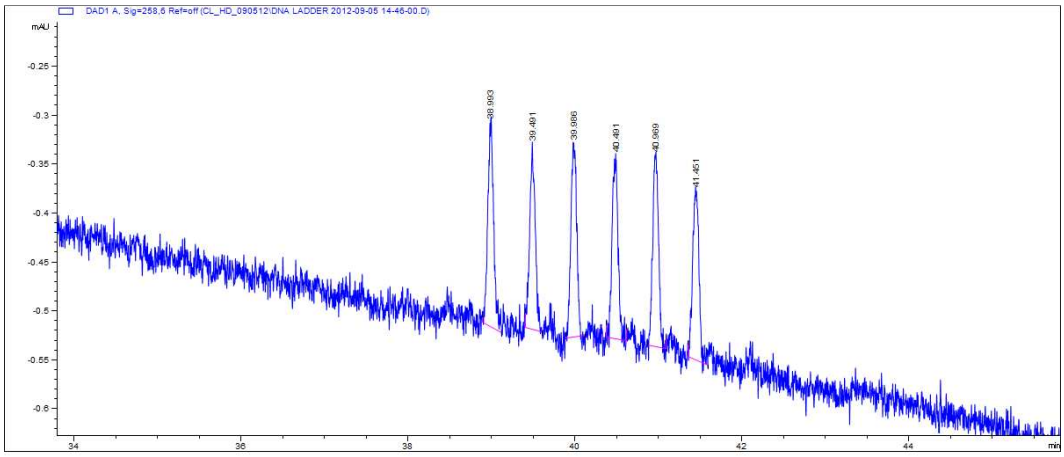


Figure 52: CGE result of the DNA ladder sample has six peaks as expected. Each of the peaks corresponds to DNA strands of length 25, 26, 27, 28, 29 and 30 bases.

The single base resolution enables the use of CGE to estimate the fraction of unlabeled DNA strands present in the fluorophore-conjugated DNA sample. Since the unlabeled strands are of lower weight than their labeled counterparts, the unlabeled strands are expected to appear as a separate peak with a shorter elution time than the peak corresponding to the conjugated strands. The fluorophore-conjugated strands used to fabricate the RET keys are diluted to 1 μM in 1x TAE Mg^{+2} buffer and run through the capillary. The CGE results of Arm 6.2 conjugated with AF 488, AF 594 and AF 647 are shown in Figures S2-S4 respectively. We observe a single dominant peak in all three cases, which indicates a negligible number of unlabeled strands in the sample.

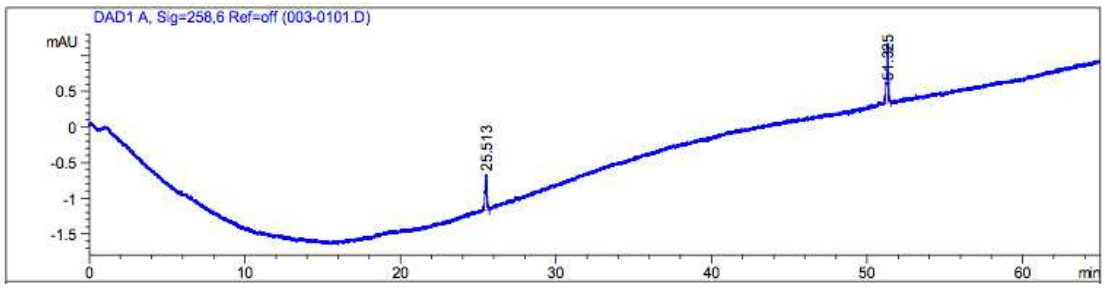


Figure 53: CGE result of Arm 6.2 conjugated with AF 488. The single peak at 51.325 minutes indicates the high yield of Arm 6.2 conjugated with AF 488. The absence of additional peaks in the vicinity of this peak indicates that the contribution from unlabeled Arm 6.2 is negligible. The peak at 25.513 minutes is due to the TAE Mg^{+2} buffer.

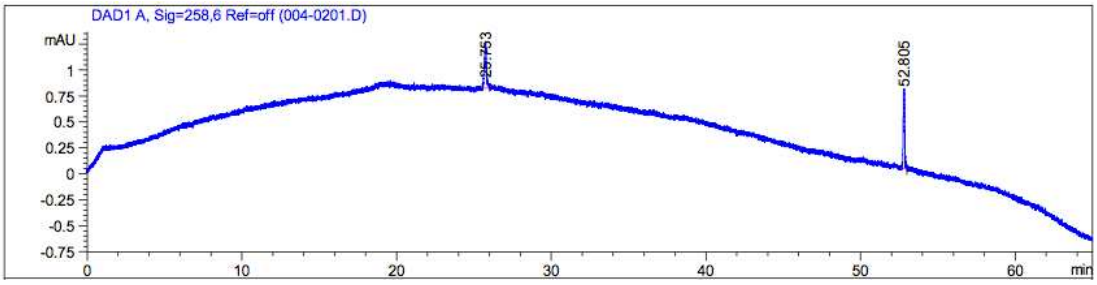


Figure 54: CGE result of Arm 6.2 conjugated with AF 594. The single peak at 52.805 minutes indicates the high yield of Arm 6.2 conjugated with AF 594. The absence of additional peaks in the vicinity of this peak indicates that the contribution from unlabeled Arm 6.2 is negligible. The peak at 25.753 minutes is due to the TAE Mg^{+2} buffer.

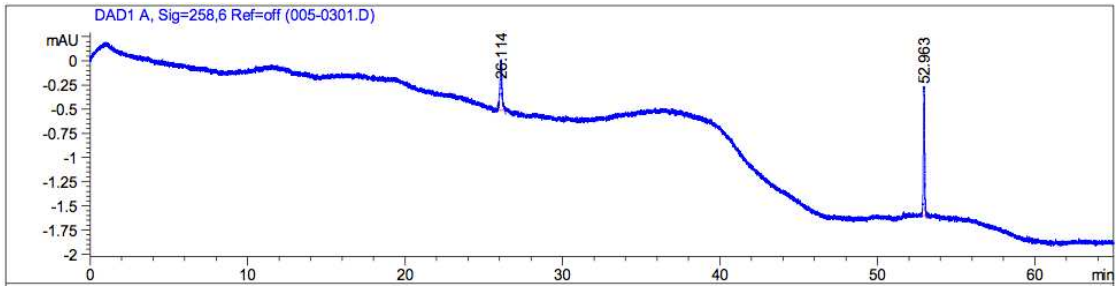


Figure 55: CGE result of Arm 6.2 conjugated with AF 647. The single peak at 52.963 minutes indicates the high yield of Arm 6.2 conjugated with AF 647. The absence of additional peaks in the vicinity of this peak indicates that the contribution from unlabeled Arm 6.2 is negligible. The peak at 26.114 minutes is due to the TAE Mg^{+2} buffer.

8.3. DNA tile and grid formation

The DNA strands are suspended in 1x TAE Mg^{+2} at 30 μM concentration. Each DNA tile structure is formed by mixing stoichiometric amounts of 1 core, 4 shell and 4 arm strands with 1x TAE Mg^{+2} (tris acetate, 40 mM, pH 8.0), EDTA of 2 mM concentration with 2 mM magnesium acetate. The volume of the strands, buffer and double distilled H_20 are adjusted to achieve a final concentration of 1 μM for each tile. The mixture is heated to 96 °C and cooled to 4 °C in 4 hours. The as formed tiles are mixed in stoichiometric amounts and annealed at 23 °C for 4 hours to form the 4-tile grid structure.

8.4. DNA structural yield quantization

Atomic Force Microscope (AFM) is used to characterize and determine the yield of DNA grid structures. The AFM images are obtained on Agilent PicoLE equipped with OTR-8 (Veeco) tips. Three microliters of the sample is deposited onto freshly cleaved mica coated with 15 μl 1x TAE, Mg^{+2} imaging buffer. After 3 minutes, 400 μl additional buffer is added to form a liquid cell around the sample and imaging is carried out in the tapping mode. Figure S5 shows the relatively large number of intact 4-tile DNA grids formed by the process described above.

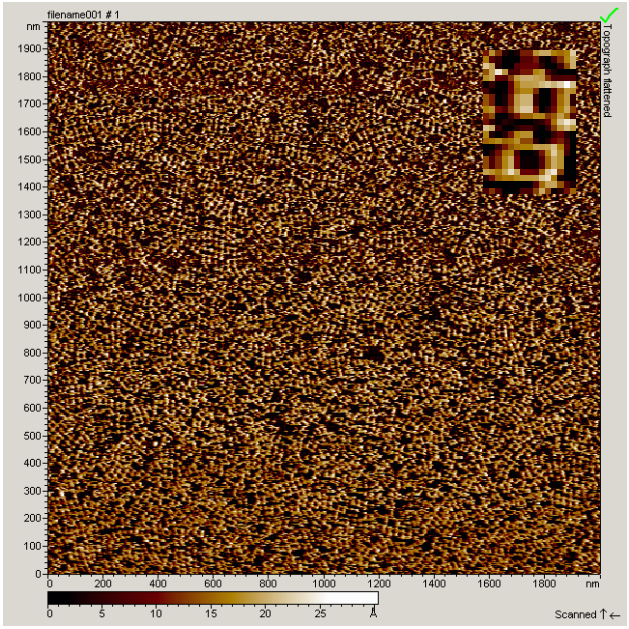


Figure 56: AFM image of a 4-tile DNA grid, indicating good yield of the self-assembled DNA structures. The inset is a zoomed in image of two 4-tile grids.

8.5. Variation of signature with keys

In the first experiment, we characterized all permutations of a three-fluorophore key by placing three fluorophores (AF 488, AF 594 and AF 647) at only four of the 6,272 available positions on a DNA grid. Fluorophores were attached to the 5′ end of SA4 in Tile 5, 3′ end of SA1 in Tile 6, 3′ end of SA2 in Tile 6 and 5′ end of Arm 6.2 in Tile 6 as shown in Figure 57. The pairwise distance between the sites where the fluorophores were conjugated is shown in Table 7 and the R_0 values between the fluorophore pairs used in the RET networks are given in Table 8. It should be emphasized here that we use

all combinations of the above-mentioned 3-fluorophore key only to demonstrate, in a reasonable amount of time, that small variations made to the key may result in a significant variation of the optical response. However, a 3-flurophore key is relatively easy to reverse-engineer using modeling attacks and therefore should not be used for practical authentication and communication. For a practical RET-key, the security increases as the size of the network increases under the constraints discussed in Section 8.6. Unlike some of the existing PUF's where increasing the complexity of the underlying structure results in higher instability of the PUF output (Herder, Yu et al. 2014), the output of the RET network is driven by an exciton mixing process, which is highly repeatable when the excitation and observation conditions don't vary as will be demonstrated later in the text.

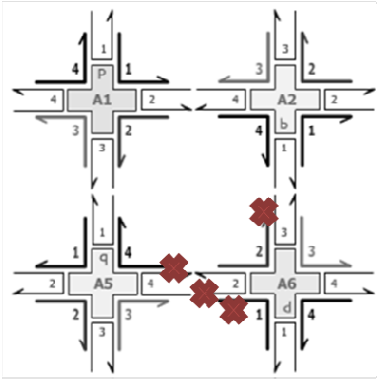


Figure 57: The figure shows the 4 sites on the DNA grid to which fluorophores are conjugated for the survey of variation in signature with minor variation to the RET network.

Table 7: Pairwise distance in nanometers between the sites where the fluorophores are conjugated for the survey of variation in signature with minor variation in the RET network.

	SA4 5' end	SA1 3'end	Arm 6.2 5' end	SA2 3' end
SA4 5' end	-	2.85	2	12.74
SA1 3' end	2.85	-	4.49	12.72
Arm 6.2 5' end	2	4.49	-	11.62
SA2 3' end	12.74	12.72	11.62	-

Table 8: The R_0 values between the fluorophore pairs used in our RET networks.

	AF 488	AF 594	AF 647
AF 488	4.87E-09	5.89E-09	5.42E-09
AF 594	1.83E-09	5.58E-09	7.80E-09
AF 647	7.37E-10	2.29E-09	6.62E-09

In all the keys, the fluorophore that is most blue in wavelength is excited and the response is observed at 543.5 nm, 620 nm and 670 nm. Figure 58 shows the change in intra-key and inter-key correlation for 648 outputs. For similar keys, under identical excitation conditions, a Gaussian fit of the intra-key correlation distribution resulted in a correlation mean of 95.46% and small variance of 6.48. Conversely, for the dissimilar keys, under identical excitation conditions, a Gaussian fit of the inter-key correlation

distribution resulted in a correlation mean of 36.42% and a large FWHM of 25.79. When half the values are discarded, there is negligible change in the distribution indicating that the number of values chosen for the analysis is sufficient.

We used the Hamming distance between signatures of similar and dissimilar keys to calculate the total number of False Negatives and Collisions in the survey respectively. The responses from AF 488 were evaluated with the following parameters: Rho resolution: 5, scaling factor: 50, threshold: 9 and tolerance: 200. On two attempts, the False Negatives were 0.64% and Collisions were 0.21%. The responses from AF 594 were evaluated using the following parameters: Rho resolution: 5, scaling factor: 100, threshold: 9 and tolerance: 400. On two attempts, the False Negatives were 0.34% and Collisions were 0.51%. Lastly, the responses from AF 647 were correlated using the following parameters: Rho resolution: 5, scaling factor: 100, threshold: 9 and tolerance: 500. On two attempts, the False Negatives were 0.94% and Collisions were 0.89%. Therefore, we see that it is possible to lower the fraction of False Negatives and Collisions to negligibly small values. A small number of False Negatives indicates that our measurements are highly reproducible while a small number of Collisions indicates high entropy in the information content of the RET-keys. However, this result was tested only on keys with a small number of fluorophores. In the next section, we create a mathematical model of large fluorophore RET-keys to analyze how the entropy scales with the size of the key.

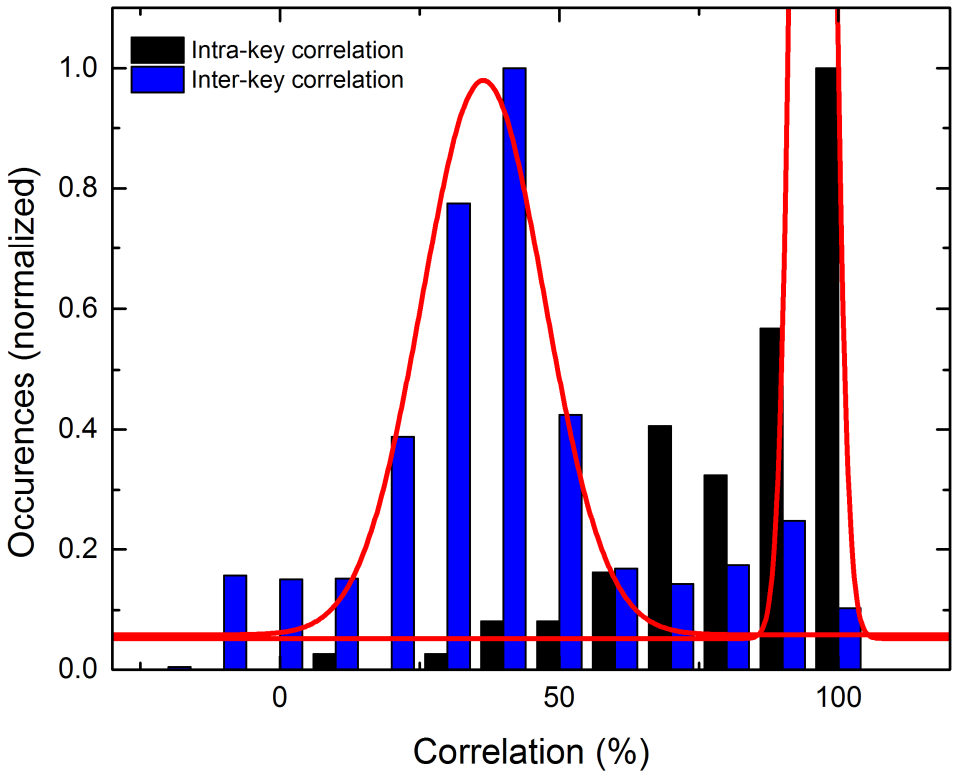


Figure 58: Intra-key and Inter-key correlation distributions for a 3-fluorophore key. It is clear that similar keys are highly correlated and have a narrow distribution while dissimilar keys have a wide distribution of correlations.

8.6. Entropy model of RET networks

The above correlation results demonstrate the avalanche effect in our keys. Avalanche effect is often used in cryptography to indicate that 50% of the output bits change when a single bit in the input is flipped. It is used more generally to indicate that

small changes in the input/key result in substantially altered output signatures. To verify whether this avalanche property is preserved in larger fluorophore networks, we performed an entropy analysis on networks ranging from 1 to 500 fluorophores.

The entropy of a network is frequently used to determine the information content of a network and is calculated using $H = -\sum_{i=1}^{n} p_i log_2 p_i$, where p_i is the weight of each edge and n is the total number of edges in the network (Cover 1991). For our analysis, we assume that the edges in the network are the transfer efficiencies between the fluorophores.

We first calculate the maximum entropy for a given number of edges in the RET network. A network is said to have maximum entropy when all the edges in the graph have equal weights, that is, all the paths in the network are uniformly distributed. The maximum entropy of a network is calculated using the formula $H = \sum_{k=1}^{n-1} u_k h_k$, where u_k is the stationary distribution and h_k is the entropy of node k. Stationary distribution is not defined for networks with absorbing as states. In the RET networks we model fluorescence as an absorbing state through which an exciton generated as per the initial condition may leave the network. Therefore, in order to compute the maximum entropy of the RET networks, we partition the transition probabilities as shown below (Saerens M 2009):

$$P = \begin{matrix} \\ Transient\ states \\ Absorbing\ states \end{matrix} \begin{matrix} \begin{pmatrix} Transient\ states & Absorbing\ states \end{pmatrix} \\ \begin{pmatrix} Q & R \\ 0 & 1 \end{pmatrix} \end{matrix}$$

Here Q is the substochastic matrix that includes the transition probabilities between transient states, R represents the transition probabilities from the transient to the absorbing states, the '0' matrix indicates that there can be no transitions from the absorbing to the transient states and the '1' vector indicates that once an exciton enters an absorbing state, it remains in that state. The maximum entropy can now be calculated using the formula:

$$H = \sum_{k=1}^{n-1} n_k h_k,$$

where n_k is the expected number of visits to each node. It is calculated using $n_k = \sum_{t=0}^{\infty}((Q^T)^t e_1) = (I - Q^T)^{-1} e_1$, where e_1 is the expected number of visits to each transient state when starting in state 1. The entropy of node k, h_k, is calculated using $h = -\sum_{k' \in S(k)} p_{kk'} log_2 p_{kk'}$, where $p_{kk'}$ is the weight of the edge between nodes k and k' and $S(k)$ is the set of all neighboring nodes from node k. For maximum entropy, every node in the network has an equal number of edges, say m. Therefore, $p_{kk'}$ is $1/m$. The implementation of the maximum entropy algorithm may be found in Appendix D. We found that the maximum entropy varies significantly with the number of edges connected to each node as shown in Figure 59. As the size of the interaction of each fluorophore increases, the size of the graph where the maximum entropy saturates also

increases. This result agrees with our intuition for RET networks. As the interaction size of a given fluorophore increases, more paths become available for an exciton to transfer to, leading to larger uncertainty in the path chosen, which in turn leads to higher entropy.

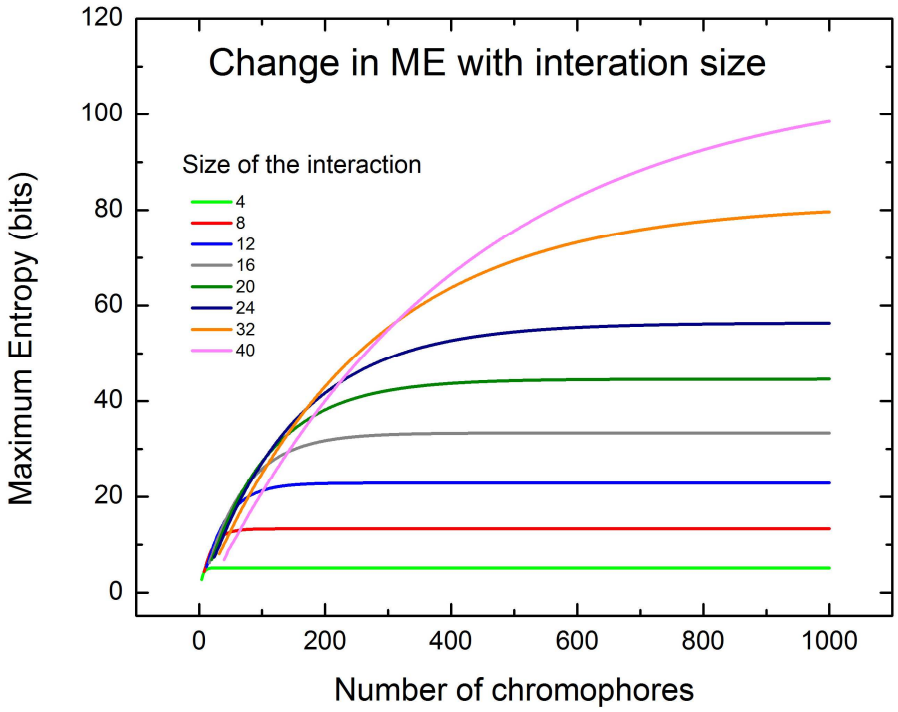


Figure 59: Variation of Maximum Entropy with the number of fluorophores in the network. We found that the interaction size of each fluorophore governs the graph size at which the maximum entropy saturates.

For our analysis, we assume that 13 fluorophores may interact with each other simultaneously. In order to calculate the interaction size, we calculated the average R_0 between the fluorophore pairs we use in the RET networks as 5.02 nm. Assuming that

we can place a fluorophore every 4 bases without compromising the yield of the structure, we should be able to place 13 fluorophores in a 25 sq. nm area. Therefore, we assume an interaction size of 13 for the maximum entropy calculation. Furthermore, each 13-fluorophore cluster can interact with other overlapping 13-fluorophore clusters.

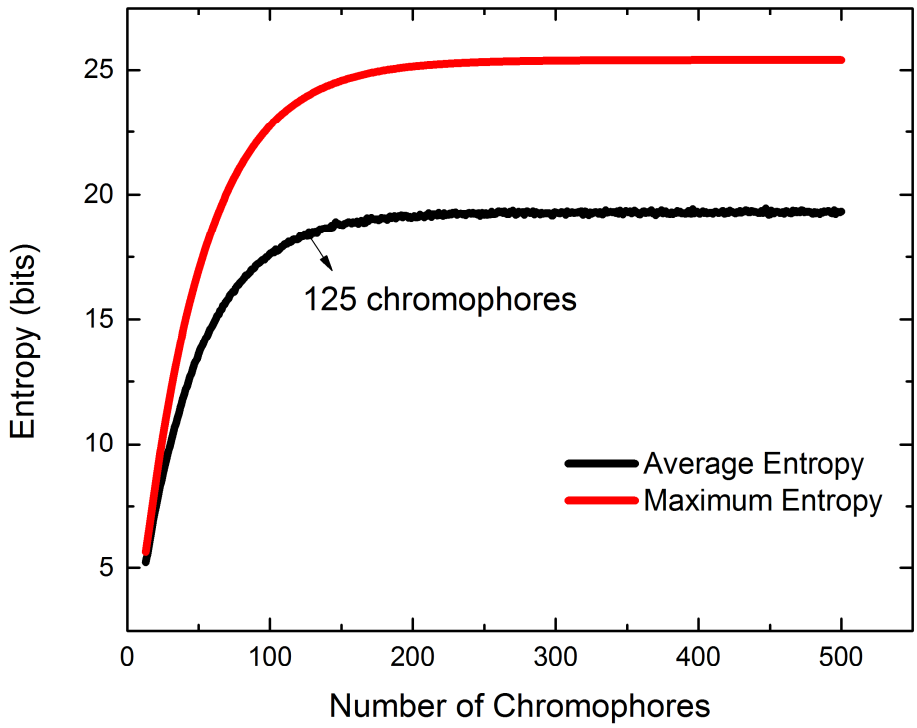


Figure 60: The figure shows how entropy varies with number of fluorophores. It is evident that after 280 fluorophores, we begin to see redundancy in the information obtained from the network.

The maximum entropy calculation only shows the entropy of one of all possible networks of a given size. In order to eliminate the bias in the calculation of entropy, we perform a Monte Carlo analysis where 500 graphs of each network size are generated and the average entropy of the 500 graphs is calculated. The interaction size in all the networks is set to 13 for the same reason described above. Again, each 13-flurophore cluster can interact with overlapping 13-flurophore clusters. Transfer efficiencies were assigned to the edges in each network randomly ensuring that all the outgoing edge weights sum to 1. The entropy of each network is calculated using $H = -\sum_{i=1}^{n} p_i log_2 p_i$, as described earlier. The implementation of the average entropy algorithm may be found in Appendix E. From Figure 60, we notice that the average entropy continues to increase up to 125 fluorophores, which we take as the maximum network size that can be used for the keys in order to minimize collisions in the output signatures. As shown in Figure 61, the value of average entropy remained the same when half the values were discarded (i.e., 250 networks of each network size are used). This result proves that the number of samples used for the calculation of average entropy is sufficient.

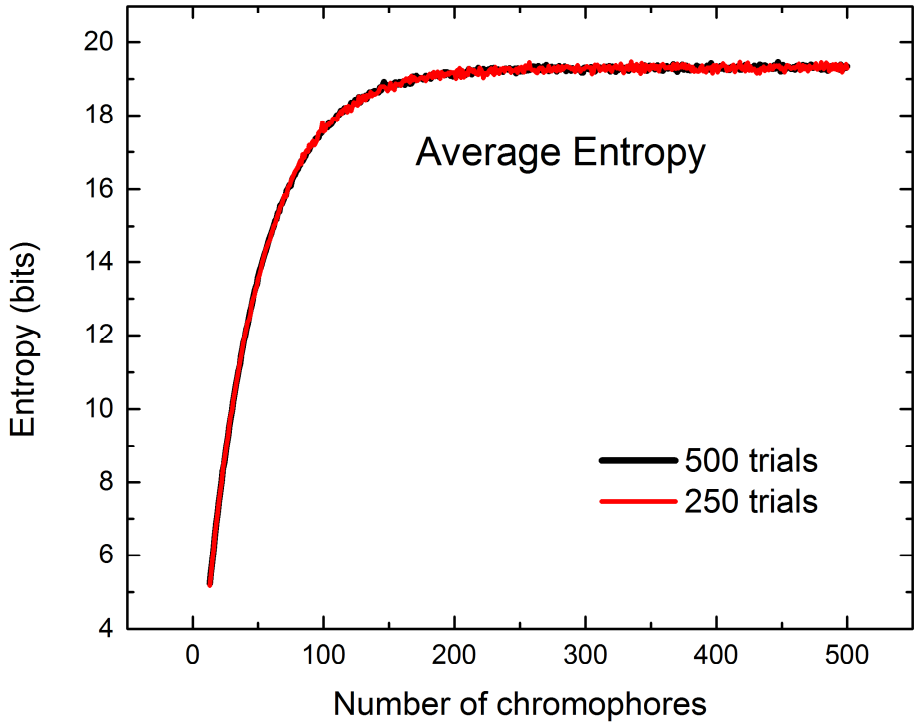


Figure 61: Variation of Average Entropy with network size. We observe that on sampling only 250 networks as opposed to 500, the entropy values remain the same. This indicates that the number of networks chosen for the calculation of average entropy is sufficient.

8.7. Variation of signature with excitation wavelength

In the second experiment, we observe the response of a single key while varying the excitation wavelength. The key contained three fluorophores, AF 488, AF 594 and AF 647 at positions shown in Figure 62.

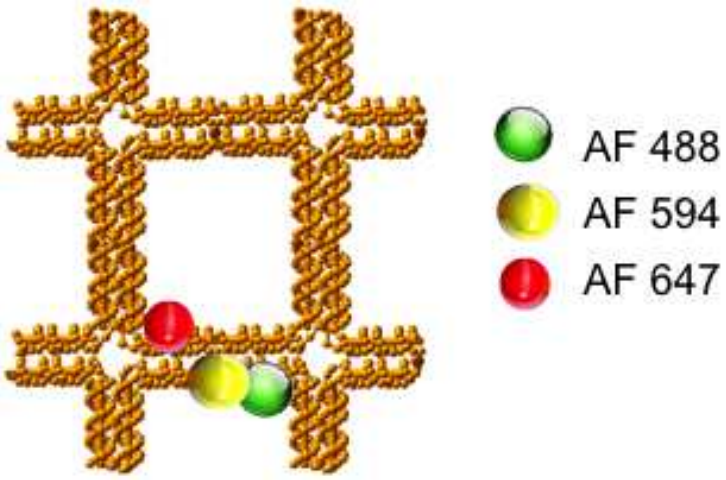


Figure 62: The RET network with fluorophores AF 488 and AF 594 placed at something distance with respect to each other while AF 594 and AF 647 placed at something distance with respect to each other. The size of the fluorophores is exaggerated in this figure for clarity.

The key is excited at the following wavelengths: 441.6 nm, 457.9 nm, 460 nm, 488 nm, 508.5 nm, 514.5 nm, 532 nm, 543.5 nm, 620 nm, 632.8 nm, 635 nm, 647.1 nm, 670 nm, 694.3 nm and 780 nm and observed at 670 nm. Under similar excitation conditions, for the same key, a Gaussian fit of the distribution resulted in a correlation mean 95.40% of and small FWHM of 5.1%. Under different excitation conditions, for the same key, a Gaussian fit of the distribution resulted in a correlation mean -28.18% of and a large FWHM of 32.04%. The smallest excitation wavelength difference that could be resolved was 2.1 nm. This was observed between the signatures of 457.9 and 460 nm as shown in Figure 64. This result indicates that we can discretize the excitation wavelength in steps

of 2.1 from 400 nm to 800 nm, which gives us a total of 190 distinct excitation wavelengths.

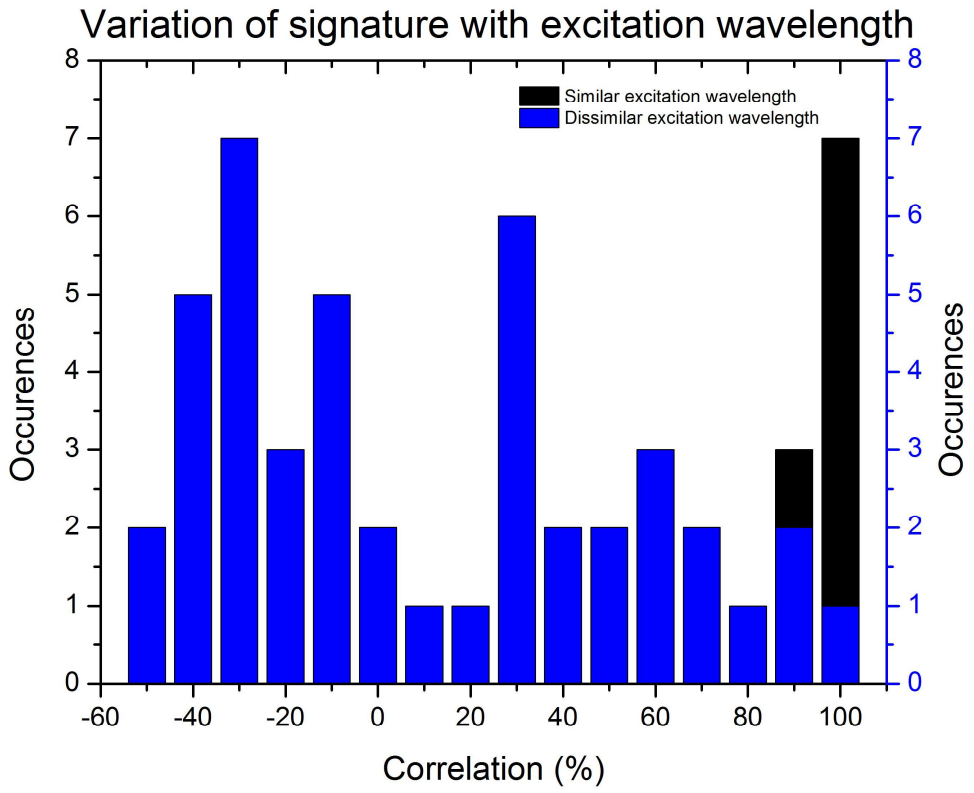


Figure 63: Intra-key and Inter-key correlation distributions for a single key but with varying excitation wavelengths. It is clear that similar keys are highly correlated and have a narrow distribution while dissimilar keys have a wide distribution of correlations.

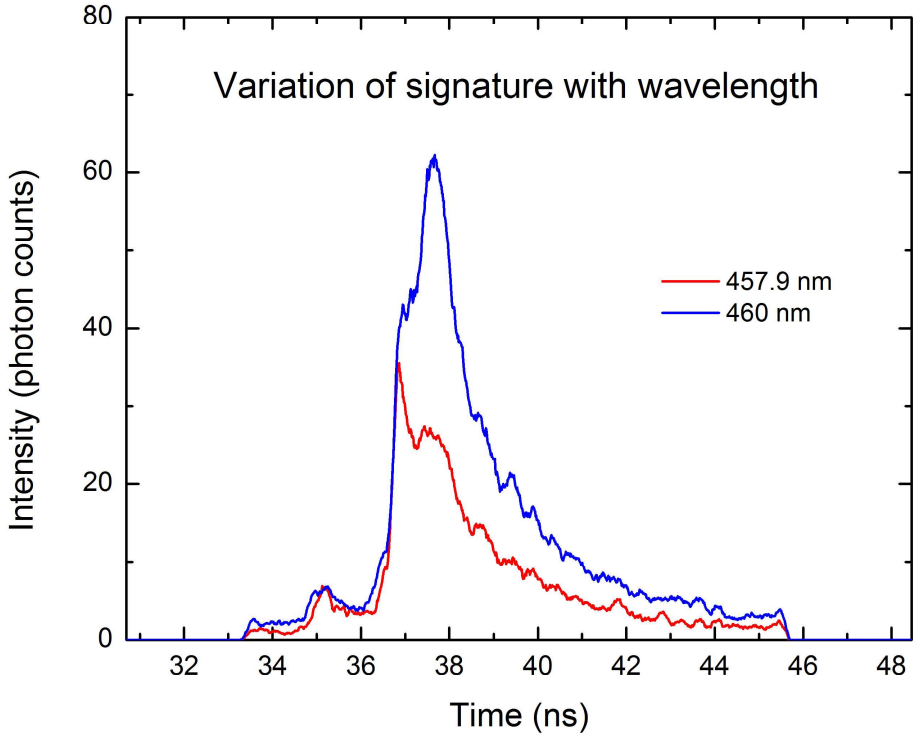


Figure 64: The output signatures of the key shown in Figure 62 when excited at 457.9 (red) and 460 nm (blue) and observed at 670 nm. From this figure, we conclude that the smallest resolvable wavelength difference that could be detected, over noise, on TREX is 2.1 nm.

8.8. Variation of signature with excitation delay

In the third experiment, we observed the response of a single key when excited at 488 nm, 543.5 nm and 647 nm with varying excitation delay. Prior to observing the variation in the signature of a single key with excitation delay, we carried out an

extensive study using the five keys shown below for over 50 excitation delays

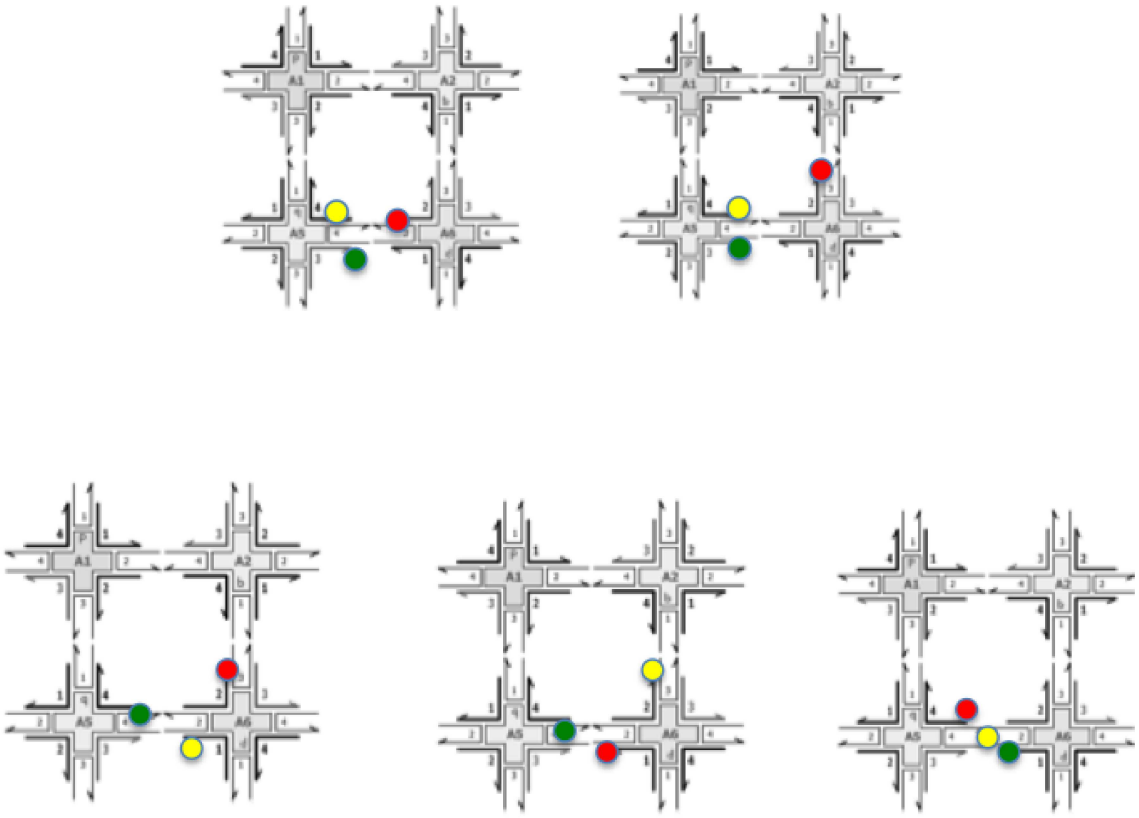


Figure 65: The responses to over 50 excitation delays on the 5 keys shown here were studied to check for any trends in the variation of the response with excitation delay. The networks were chosen such that the responses covered a wide range of lifetime and amplitude values. The green full circle indicates AF 488, the yellow circle indicates AF 594 and the red circle indicates AF 647.

The conclusions from this study are listed below:

- Correlations were calculated for signatures obtained using the filters 543.5 nm, 620 nm and 670 nm, which are the emission wavelengths of AF 488, AF 594 and AF 647 respectively. We found that the signatures obtained using the 620 nm filter offered the highest separation between the intra-key and the inter-key correlations. This is due to the widely varying inter-key correlations caused by the large lifetime range of AF 594 compared to the relatively narrow lifetime range of AF 647.
- Highest separation between the intra-key and inter-key correlations was seen when the fluorophore being observed was not excited. The coupling between the excitation wavelength and the emission filter results in reduced discrimination between of the output signatures even for different excitation conditions.
- Higher separation was noticed between the intra-key and inter-key correlations when all three excitation pulses were separated in time compared to two pulses separated in time, which was better than having all three pulses arriving at the at the same time. This is again due to improved inter-key correlations since the introduction of delays between pulses enables the effect of saturation to manifest in the output signature.
- On repeating measurements with the same key and identical excitation conditions, we obtained high agreement in the correlations with the

measurements on identical samples under identical excitation delays taken previously.

We considered a single key containing three fluorophores, AF 488, AF 594 and AF 647. The positions of the fluorophores are shown in Figure 62. The keys are excited at 488 nm, 543.5 nm and 647 nm and observed at 620 nm and 670 nm. The excitation delays varied from 0-4 ns in steps of 500 ps. Under similar excitation conditions, for the same key, a Gaussian fit of the distribution resulted in a correlation mean 94.85% of and small FWHM of 0.05%. Under different excitation conditions, for the same key, a Gaussian fit of the distribution resulted in a correlation mean 35.5% of and a large FWHM of 57.34%.

In order to determine the smallest delay that may be resolved using our instrument, we carried out a similar experiment as the one described above but with the excitation delays of the 543.5 nm and the 647 nm laser pulses varying from 0-1 ns in steps of 100 ps with respect to the 488 nm pulse. As shown in Figure 66, a time offset of 100 ps resulted in a clear discrimination between the signatures of similar and dissimilar excitation delays for an identical key under the same excitation wavelengths.

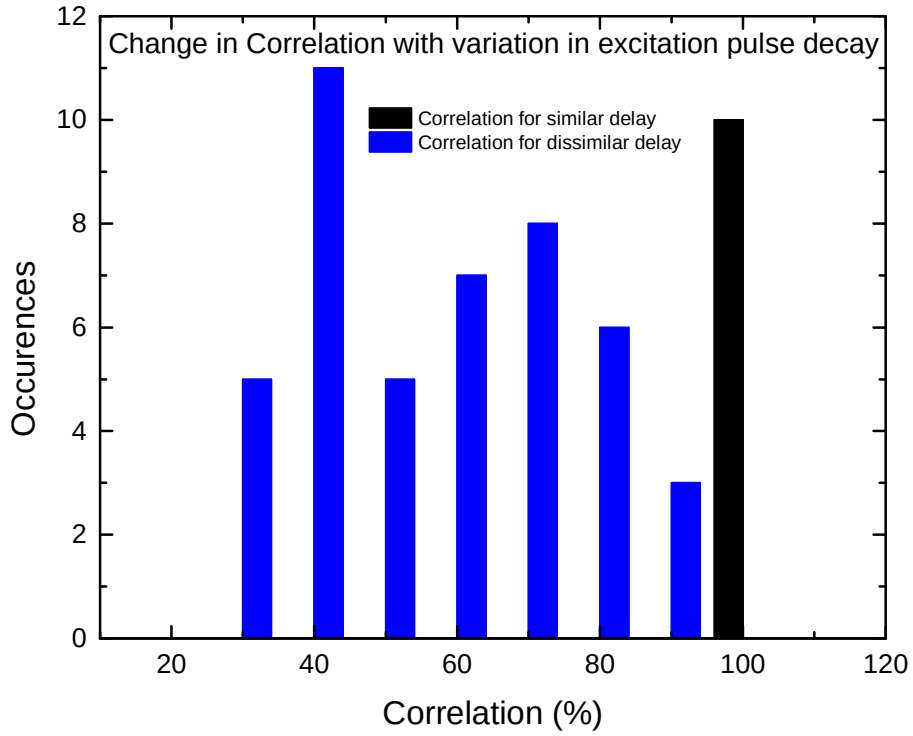


Figure 66: Intra-key and Inter-key correlation distributions for a single key and identical excitation wavelengths but with excitation delays varying from 0-1ns in steps of 100 ps. Since there is no overlap between the intra-key and inter-key distributions, we conclude that similar histograms offset in time by at least 100 ps can be resolved using our instrument and signature generation algorithm.

8.9. Repeatability test

In the repeatability study, a series of coterminous measurements are recorded on two instances of the network shown in Figure 62. The samples are excited at 488 nm for 100s and observed at 670 nm. The signatures of the two samples were compared using

the following parameters: Scaling factor =50, Threshold = 9 and Tolerance = 100. The purpose of this study is three-fold:

1. To ensure that the samples being used for authentication by different parties bleach at the same rate.
2. To experimentally verify that it is possible to get different signatures between successive measurements due to a change in concentration or bleaching.
3. To estimate the number of authentication attempts a certain sample could be used for.

We found no false negatives between the signatures of two batches of the same RET key, indicating that the samples evolved at the same rate with time. Additionally, the collision between signatures of successive measurements of the same batch is as low as 0.61% for the parameters listed above. This indicates that the sample does change with time. At the end of 15 measurements, we notice a drop in the total volume of the sample, which indicates that the concentration of the sample increased in addition to bleaching. We notice that a drop in the total counts between successive measurements is 1.44%. A sample with a total of 100,000 counts at the first measurement could then be used roughly 476 times before it reaches the minimum detectable signal of 100 counts. This rate, however, depends largely on the network, the excitation wavelength used, the excitation delays, laser intensity and temperature.

Temperature plays a crucial role in determining whether the rate of evaporation or bleaching dominates in varying the signature between successive measurements. At low temperatures, very little volume is lost and we clearly see a drop in photon counts between successive measurements on the same sample. Whereas, at higher temperatures significant evaporation takes place and as a result the increase in concentration of the sample dominates any effect bleaching may have. This results in the photon counts going up between successive measurements. Therefore, temperature of the measurement could serve as a control parameter that determines the signature of a sample. A detailed description regarding the measurement conditions required for successful authentication between legitimate users is provided in the following chapter.

In summary, we fabricated and characterized a large number of three-fluorophore RET-keys to evaluate the sensitivity of the responses from the keys to minor variations in the key or the excitation conditions. It is the most exhaustive survey of RET-keys under varying excitation/observation conditions to the best of our knowledge. From experimental results, we were able to verify that responses from similar keys under similar excitation conditions correlated very highly. These correlations were well separated from correlations between dissimilar keys under similar excitation conditions or similar keys under dissimilar excitation conditions and these correlations were seen to

vary over a wide range. We also noticed that observing the response of different keys under a few different excitation/observation conditions could result in a negligible number of collisions between the signatures of dissimilar keys. These properties could potentially be useful in the parallel detection of a large number of entities, if they can be assembled into a RET network as well as the structural detection of molecular entities. The optical nature of the excitation and response combined with the biocompatible DNA substrate could make it particularly attractive as an in situ biological sensor. Such a sensor should potentially be able to detect multiple molecular species in parallel with high sensitivity and accuracy. We will further describe the need for such a sensor and the feasibility of designing it in Chapter 11.

9. RET-key Authentication

Authentication is the process of verifying that an authorized transmitter using an established protocol, that has not been modified or tampered with, has sent a certain message. In the case of the RET-key, if Alice wants to authenticate Bob (see Scheme 1), she sends a challenge constituting the excitation wavelength, excitation delay and observation wavelength to him. In addition to the challenge, a set of helper data is transmitted to minimize the difference in the measurements between the two parties. The helper data consists of the sample volume, sample concentration, excitation time, laser intensity, total number of counts and the TCSPC settings. Bob then computes a histogram by applying the challenge on his key and sends the output to Alice. Since Alice and Bob possess identical keys, Alice checks the histogram from Bob against her own and if both sets of data match within the allowed tolerances, authentication is established.

The challenge and response strings are in the public domain but they are XOR'ed with a cryptographic obscurer derived from the key. The obscurer is the output bit string obtained independently by Alice and Bob when they apply an identical, public input to their respective keys. The large output space enabled by our scheme ensures that an adversary, not in possession of the right key, will have a 1 in 10^{375} chance at

identifying the correct obscurer. Making use of a cryptographic obscurer, therefore, strengthens our protocol against passive attacks (Goldreich 2001).

Authentication with the obscurer involves two phases: the generation phase and the reproduction phase. During the generation phase, the message to be communicated is XORed with the obscurer as shown below and the derived string is made public. During the reproduction phase, a legitimate user of the key can XOR this public string and the obscurer derived from their key to recover the original message.

Generation:

Message: 1100001111100

Obscurer: 0001111001000

Ciphertext: 1101110110100

Reproduction:

Ciphertext: 1101110110100

Obscurer: 0001111001000

Message: 1100001111100

Using the obscurer will strengthen our authentication protocol if has the following 2 properties:

1. Both the authenticating parties should arrive at precisely the same bit string.
2. The information that is made public to compute the obscurer should not reveal any information regarding the underlying RET network.

The obscurer consists of a row in the Hough matrix, that is, all the angles corresponding to a certain distance. Alternatively, we can choose a obscurer that is a random combination of shorter bit strings across many distances. We found that due to minor variations in the samples or measurement noise, some inconsistencies between the obscurer measured by the two authenticating parties are inevitable. In order to generate the correct obscurer with high precision, the Reed-Solomon or BCH error correcting codes may be employed (Guajardo, Kumar et al. 2007, Bosch, Guajardo et al. 2008).

In order to compute the obscurer, Alice and Bob publicly agree on the excitation wavelength, excitation delay and the observation wavelength. This information does not help the adversary identify the underlying network since the response of the key is private. Consequently, an adversary does not gain access to even a single challenge-response pair and therefore cannot identify a subset of keys that have an identical response to the public challenge. With regards to the helper data, the total number of counts would have to be made public in order to compute the obscurer. Making the total counts in the sample public knowledge does not disclose any information regarding the

signature of the sample when the challenge and response are obscured. For instance, a total of 10,000 counts could be distributed across a single exponential decay or as many as 25 exponential decays depending on the number of distinct excitation inputs. An adversary therefore, would have to go through all combinations of response profiles that result in a total of 10,000 counts. It might make it slightly easier to attack since an adversary only has to try a total of 10,000 counts as opposed to all possible total count values. The attack now would require $(c_1)(n-c_1)(n-c_1-c_2)\ldots\ldots(n-c_1-c_2\ldots\ldots-c_n)$, here n is the total counts and c_i is the total number of counts in the response to the i'th excitation delay. For 10,000 counts, the total number of single exponential decays that can be resolved with TREX is 60. For 125 distinct excitation pulses, the total number of obscurer values the adversary will have to choose from are 60^{125}.

9.1. RET-key Advantage

Cryptographic advantage is defined as the amount of time needed for an adversary as opposed to a legitimate user to arrive at the right key. For the RET-key, this quantity is the product of the total number outputs possible and the time it takes to try each output. In attempt to realistically determine the total number of outputs, we calculate this quantity by taking the current instrument resolution and experimental noise into account. We know that TREX can detect ≈200,000 peak counts without

discarding additional photons. The minimum signal that we can hope to resolve distinctly over noise, using TREX, is around 100 peak counts. We noticed that we could reliably detect an 8% change in amplitude in the above-mentioned range. Therefore, any signature falling within 8% of another is considered the same. The range of lifetimes that we can detect with the fluorophores we use is roughly 0.1 ns – 6 ns. We noticed that we could resolve a 14% change in lifetime using TREX. Therefore the total number of unique single exponential decays that we can resolve is 100 amplitudes x 20 lifetimes. This gives us a total of $2x10^3$ unique signatures for any given combination of excitations and keys. The set-up can therefore detect $2x10^3$ unique signatures for a single exponential decay. Since a 100 ps delay gives rise to a new experimentally detectable exponential decay, a total of 125 excitation decays can be resolved using our instrument. Therefore the total number of outputs that can be measured using our instrument is $(2x10^3)^{125}$ or $\approx 10^{375}$ unique signatures.

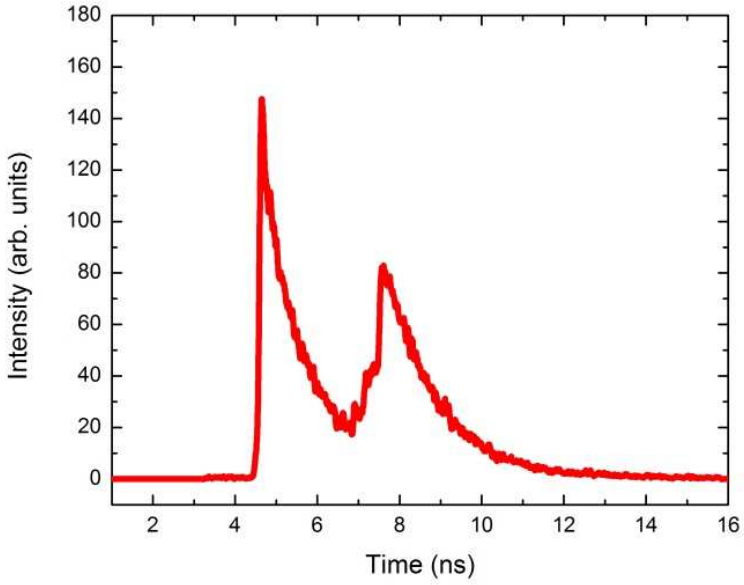


Figure 67: Sample output with two excitation delays. The use of multiple excitation delays reduces the number of collisions between different excitation-key combinations.

The use of multiple excitation delays has a critical advantage of reducing the number of collisions between different input/key combinations. The use of multiple excitation pulses, offset in time, enables the effect of fluorophore saturation to manifest in the time-resolved histogram. Fluorophore saturation refers to the fluorophore being in the excited state due to which none of the other fluorophores in the network can transfer to it. This creates a different transfer network from the one in which saturation effects are not considered, resulting in a larger range of time-resolved fluorescence histograms from the RET keys. Changing the observation wavelength also results in a variation of signatures despite the same input/key combination. From the survey of RET keys, based on the observation wavelength, the samples that collide change as shown in

Figure 68. We found that of the samples that collide at an observation wavelength of 543.5 nm, 11.3% also collide at an observation wavelength of 620 nm. Of the samples that collide at an observation wavelength of 620 nm, 11.12% collide at an observation wavelength of 670 nm. Of the samples that collide at an observation wavelength of 543.5 nm, 6.33% collide at an observation wavelength of 670 nm. Therefore, despite noticing a collision between two excitation-key-observation combinations for a particular wavelength, it is highly unlikely for two keys to collide under all excitation/observation conditions.

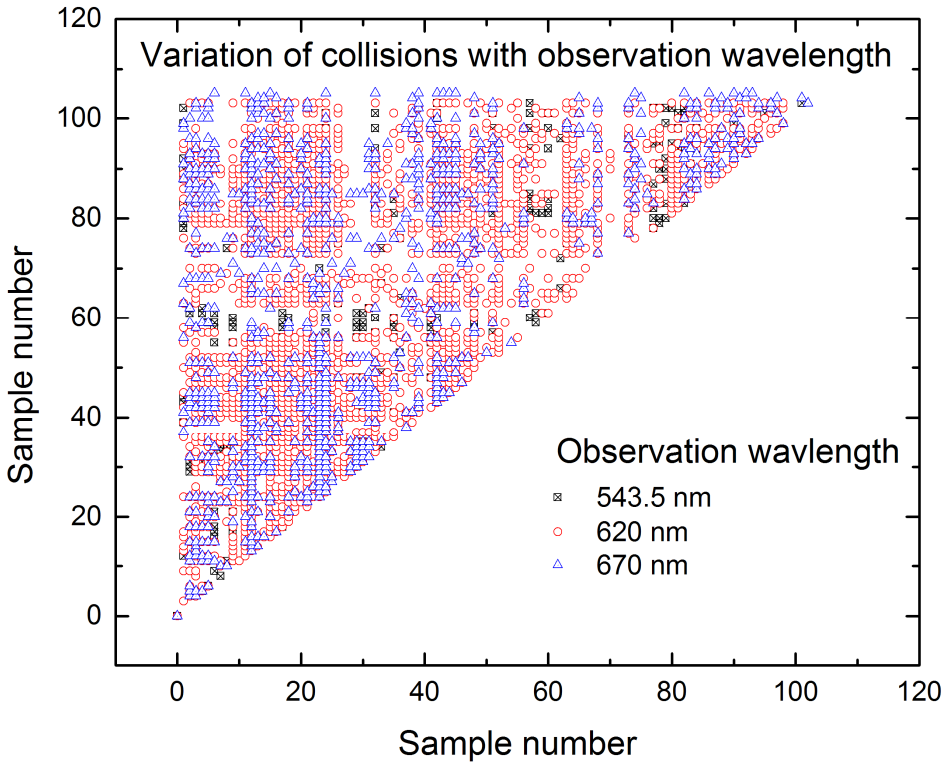


Figure 68: The figure shows the collisions in the output space when the observation wavelength is 543.5 nm (black square), 620 nm (red circle) and 670 nm (blue triangle). We notice that the collisions vary significantly with the observation wavelength for the same key under identical excitation conditions. Therefore, it is possible to significantly lower the collisions in the responses by using multiple excitation pulses are used with a time offset instead of a single excitation pulse.

The use of multiple excitation channels, therefore, enables us to realize a computation advantage of 10^{375} over an attacker. This is significantly higher than the advantage shown in other keys listed in Table 9. It is important to note here that Pappu et al.'s work is impractical to break using current computational resources. Our claim

here is that the RET-key is impractical to break as well due to the higher cryptographic advantage. Furthermore, an attacker cannot try only a subset of the 10^{375} outputs since they do not have access to the challenge used and do not have information about the underlying RET network. The challenge between two users is XOR'ed with a cryptographic obscurer and hence the attacker cannot guess how many input pulses are applied and at what time delays, which means they have to try all combinations of up to 125 excitation pulses. Additionally, since the attacker does not have information regarding the RET network used in a key, they cannot estimate the amplitude or lifetime information of the time-resolved output that results from each of the excitation pulses applied to the key.

Table 9: Comparison of the total number of input/key combinations of the existing keys. It is evident that the RET-key has the potential for a substantially higher advantage over other keys.

	Pappu	Lofstrom	Lee	Suh	Stanzione	Our device
Number of trials for Brute Force Attack	10^{69}	6216	$1.4*10^{20}$	523776	10^{25}	10^{375}

We will be generous to the attacker and assume he has access to a billion computers, each has 10 GHz capacity, and that each simulation only takes 1 ns. Under these assumptions, it would still take 10^{340} years for him to simulate all possible input/key combinations.

In the following chapter, we will use the advantage result of this chapter to analyze the RET-key's response to a broad range of the cryptographic attacks.

10. Cryptographic Attacks on the RET-Key

In this chapter we survey all possible attacks an adversary may use to break or reduce the strength of the RET-key. There are two broad categories of potential attacks on the key: invasive and non-invasive. Invasive attacks attempt to establish the structure of the key by physically examining it. Most keys are built so as to visibly change the properties of the key once it has been through an invasive attack. In the case of the RET-key, we have shown earlier that obtaining physical access to the device will not reveal any information about the fluorophore network on the key. Non-invasive attacks attempt to identify the underlying structure of the key by observing the input, output or both. We considered the RET-key's response to all available non-invasive attacks and the ones with the highest probability of success are detailed below. In order to analyze the strength of the key against various cryptographic attacks, we assume at all times that the adversary has complete knowledge of DNA self-assembly and RET networks. Therefore, information pertaining to the assembly of the DNA constructs, positions of the fluorophores and the range of input wavelengths and excitation times is public knowledge.

10.1. Brute Force Attack

In cryptography, a brute force attack is independent of the physical system and probe used. It is usually applied when all other forms of attacks are ineffective. The attack tries all possible responses and tries to estimate the right one. The adversary lacks knowledge regarding the inputs applied or the actual physical structure of the key being used and instead attempts all possible outputs of all keys. In the case of the RET-key, the adversary has to try all 10^{375} outputs. In the case of the RET-key, the brute force attack is complicated by the multiexponential nature of the fluorescence decay curves. It is not possible for an adversary to simply try all combinations of amplitudes and lifetimes because of the problem of correlation parameters in resolving multiexponential decays. In the equation below, the same intensity may be obtained for different values of amplitudes and lifetimes. If the total number of fluorophores exceeds three, resolving individual decay rates accurately is extremely difficult.

$$I(t) = \sum_{i=1}^{n} A_i e^{-\frac{t-T}{T_i}}$$

where,

A_i Amplitude of the ith component in the first fitting range channel

T_i Lifetime of the ith component

T Time delay

10.2. Birthday Attack

A birthday attack is based on the birthday problem described in probability theory and computes the probability of finding two people in a room with the same birthday given a fixed number of trials (Stinson 2006). In the case of the RET-key the problem can be restated as identifying the probability of two inputs resulting in the same output give a fixed number of trials, say k and all possible outputs, m. In our system, we attempt to identify the number of trials it takes to arrive at a collision probability of 0.5. A high collision rate does not affect a legitimate user's chances of arriving at the correct answer given an input and key pair. However, a high collision rate decreases the entropy of the system, limits the output space and reduces the number of attempts required by a brute force attack. The probability that 2 values hash to the same value is given by $e^{\frac{-k(k-1)}{2m}}$', as shown below (Stinson 2006) :

Probability of no collisions: $p = \frac{m(m-1)(m-2).....(m-k+1)}{m^k}$

Probability of at least one collision: $1 - p = \left(1 - \frac{1}{m}\right)\left(1 - \frac{2}{m}\right) \ldots \ldots \left(1 - \frac{k-1}{m}\right)$

First order Taylor expansion: $e^x = 1 + x$

$= 1 - p = e^{-\frac{1}{m}} e^{-\frac{2}{m}} \ldots \ldots e^{-\frac{k-1}{m}}$

$= e^{\frac{-(1+2+\cdots\ldots(k-1))}{m}}$

$= e^{\frac{-k(k-1)}{2m}}$

For 50% probability that 2 values hash to the same output: k =

$\sqrt{1.38x10^{375}}$
$= 1.17\ x10^{187}$

In the case of the RET-key, for a 0.5 probability of a collision, we need a total of 10^{187} attempts. It is important to note here that the 10^{187} is the total number of attempts assuming that there are 10^{375} *unique* outputs.

10.3. Cipher text attack

Cipher text indistinguishability is the property by which an adversary cannot learn any useful information about the physical system by obtaining access to multiple ciphertexts. In the RET-key, the ciphertext is the excitation wavelength and the delay of the excitation pulse. It is possible to conclude that the key is an optical one but the exact nature and range of the outputs will remain unknown. Furthermore, the challenge is made public only during the computation of the obscurer. Every subsequent challenge is XOR'ed with the obscurer and therefore is not accessible to an adversary.

10.4. Replay attack

Another possible attack on the RET-key is the replay attack. In the replay attack, the attacker stores the challenge-response pairs of all the RET-keys and eavesdrops on the communication channel used by the legitimate users in order to identify the key being used. Each challenge of the RET-key requires 1875 bits: A user can specify at most 125 excitation pulses, each of which requires 8 bits to specify a wavelength between 350 nm and 800 nm in steps of 2.1 nm and 7 bits to specify a delay between 100 ps and 12,500 ps in steps of 100 ps. Each response of the RET-key is 900 bits and is derived from the Hough transform matrix. Therefore, each key requires roughly 2775 bits to specify the input and output and since there are 10^{375} input-key combinations, we would require roughly 346×10^{375} bytes in order to store the challenge response pairs of all keys. Furthermore, the response of the key to the same input will be different in time as shown earlier. The total CRP space, if this time dependence is taken into account, will be much larger and therefore more than 10^{375} bytes of memory will be required. The world had the capacity to store only 10^{20} bytes in 2007 according to (Hilbert and López 2011), which makes the replay attack infeasible using the RET-key.

While using the RET-key for authentication, it is possible that an attacker intercepts a communication channel between Alice and Bob, acquires a number of challenge response pairs and uses them in a replay attack. Since the response of the RET

key is time dependent, the response of the key recorded for a particular challenge will be different when used at a later time. Furthermore, an adversary cannot gain access to the true challenge-response pairs since they are encrypted with the cryptographic obscurer, known only to the legitimate users of the key.

10.5. Man in the middle attack

In order to prevent a man in the middle attack, where an adversary may replace the original key before it reaches the legitimate user, Alice and Bob authenticate their keys with the manufacturer upon arrival. Since the key need not be transported during subsequent stages of communication, a one-time authentication with the manufacturer should suffice. The manufacturer need not retain a copy of a RET-key once all the legitimate users authenticate their keys. A related attack is a dilution attack, where an adversary steals a negligible part of the key, dilutes it and attempts to scale the signature to match the original concentration of the key. Limiting the volume of the material exchanged such that the loss of a single DNA grid can be detected can prevent the dilution attack.

10.6. Side Channel attack

A side channel attack is where an adversary uses the information obtained regarding the manufacturing process or the physical implementation of the key in order to decipher the underlying structure of the key. Such an attack is possible when the adversary uses social engineering or coercion to extract the required information from the manufacturer. Alternatively, the manufacturer may be untrustworthy. For the RET-key, the manufacturer is the trust anchor and the security of the manufacturing facility is assumed to be uncompromised at all times.

11. Multiplexed Fluorescence Sensor for Cancer Detection

Through the extensive study of three fluorophore RET networks and the mathematical modeling of larger fluorophore networks, we have demonstrated that a large number of experimentally resolvable signatures are possible even with a small number of fluorophores. These signatures are extremely sensitive to small changes in the network and yet are highly reproducible. We propose using these two results to design a RET network based fluorescence sensor that can capture molecular differences in DNA/RNA secondary and tertiary structure at picosecond time resolution.

11.1. Motivation

Fluorescence microscopy has been used extensively to study biological systems. However, the number of spectrally distinguishable fluorophores limits the multiplexing ability of the technique. The number of distinct sub-micrometer fluorescence barcodes that have been demonstrated in situ is only 11 (Lin, Jungmann et al. 2012). Geometrically encoded barcodes are capable of generating much larger barcode libraries and have been demonstrated in situ(Lin, Jungmann et al. 2012). Recently, it was shown that super resolution imaging of geometrically encoded fluorescence barcodes could detect 216 barcodes uniquely, in parallel. These barcodes consists of fluorophores placed on a self-assembled DNA structure. The authors use DNA-PAINT to record super-resolution

images of the barcodes. As shown in Figure 69, this technique enabled the detection of fluorophores placed below the diffraction limit of light.

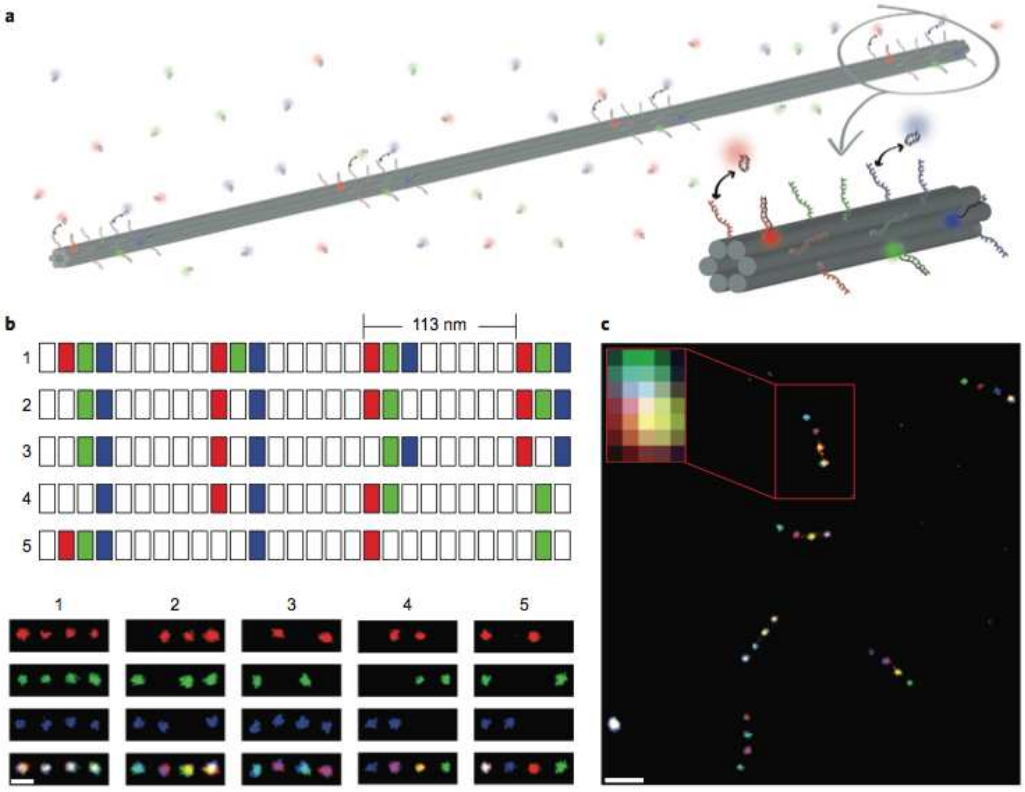


Figure 69: Geometrically encoded fluorescence barcodes can detect 216 barcodes in parallel using super resolution imaging. Figure from (Lin, Jungmann et al. 2012).

While this work established a new standard in the number of barcodes that can be detected uniquely in addition to significantly increasing the spatial information density, it has several disadvantages:

1. The nanostructure is 400-800 nm long since triplets of fluorophores have to be separated by over 100 nm for super-resolution imaging. Since a cell is a few microns in size, it limits the total number of barcodes that may be placed inside a cell. Crowding of barcodes within a cell may make it difficult to accurately resolve individual barcodes.
2. The number of spectrally resolvable fluorophores remains low using this technique and as such, geometrical encoding and imaging the barcodes becomes necessary.
3. Super-resolution imaging is extremely expensive and time-consuming.

The above limitations can be overcome by using DNA/RNA self-assembled RET networks. Fluorophores attached to single stranded DNA/RNA can self-assemble *in vivo* to form a nanoscale optical network. On exciting the donor fluorophore with an appropriate wavelength, energy is transferred to the acceptor fluorophores and subsequently released as fluorescence. The fluorescence is then observed in time at picosecond resolution. From entropy modeling of the RET networks, we know that the use of time-resolved fluorescence enables the observation of 125 fluorophores

simultaneously. Furthermore, entropy modeling and experimental analyses of over 1200 time-resolved fluorescence signatures on 108 unique networks show that the optical responses are highly repeatable and minor variations between RET networks can be discriminated resulting in a total of 10^{375} unique responses. The use of RET networks as a biological sensor offers several advantages over existing in situ sensors:

1. Ideally, the fluorophores in the RET network should be placed within a few nm of each other, which makes the size of the network extremely small.
2. The small size of the network and the extremely high spatial information density enables its use as a subcellular probe.
3. The use of time-resolved fluorescence detection ensures that a large number of signatures can be uniquely detected.
4. The use of time-resolved fluorescence detection enables the detection of small variations in the position of fluorophores.
5. The characterization technique used is TCSPC, which is significantly faster and substantially lower in cost compared to super-resolution imaging. Additionally, since we don't rely on imaging, the sensors we use can potentially be used to study live cells. Such an optical sensing mechanism enables the *in vitro* and *in vivo* characterization of the structure at picosecond resolution.
6. The small size of the oligonucleotides and fluorophores results in molecular scale spatial resolution.

The RET network can be used to monitor ligand binding, protein binding to RNA, binding of small regulatory RNA molecules to mRNA, monitor changes in secondary or tertiary structure due to mutations in the DNA/RNA sequence, changes in pH, changes in ionic conditions and temperature variations (Lakowicz 1999). In this work, we will use the increased spatial information density, molecular scale spatial resolution, high specificity and small size of the RET network to identify molecular scale structural differences between cancerous and normal cells. This can help in the detection of cancer and may also aid in the development of highly specific therapies for cancer.

11.2. Detection of cancer cells using the RET network

The knowledge of DNA and RNA secondary and tertiary structure is extremely important in understanding their function (Tucker and Breaker 2005, Wan, Kertesz et al. 2011, Dillon, Pierce et al. 2013, Wan, Qu et al. 2014) (Kertesz, Iovino et al. 2007, Zhao 2010, Rouskin 2013, Wan, Qu et al. 2013). The secondary and tertiary structure in the regulatory regions of genes is particularly important for the regulation of protein synthesis and has been evolutionarily conserved (Mortimer, Kidwell et al. 2014). It is now known that the presence of aberrant DNA/RNA secondary structure in the regulatory regions of genes involved in cell proliferation, cells growth and apoptosis can lead to cancer (Stoneley and Willis 2003). The presence of secondary structures in DNA, such as G-quadruplexs, I-motif's, triplexes, cruciform motifs and slipped structures

result in formation of fragile sites in the human genome, which in turn results in chromosomal breakage that can lead to cancer (Zhao 2010). Therefore, the ability to detect the secondary and tertiary structure specific to cancer cells can be used to detect the presence of a certain type of cancer and may also treat the cancerous cell with high specificity.

Existing techniques to determine nucleotide structure include hydroxyl radical probing, divalent lead ion footprinting, dimethyl sulphate (DMS) and selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE). The advantages and disadvantages of these techniques at secondary and tertiary structure detection are briefly described here.

Hydroxyl radical footprinting relies on strand scission that takes place when the RNA backbone is modified (Shcherbakova and Brenowitz 2008). This method can be used to determine tertiary structure and can be performed with millisecond time resolution. However, since hydroxyl radical footprinting is not specific to any particular base, it is difficult to use the technique as a sensor with high specificity to detect certain sequences of DNA or RNA (Stefanie and Kevin).

A divalent lead ion has been demonstrated as an *in vivo* structural probe. Similar to hydroxyl radical probing, the lead ion cleaves unpaired RNA while paired RNA is

mostly left undisturbed (Magnus, Pascale et al. 2002). However, a synchrotron is needed in order to use this technique, which makes it impractical and extremely expensive.

In DMS footprinting, DMS donates a methyl group to the hydrogen bond accepting ring nitrogens at the N1 site of unpaired adenosine and the N3 site of cytidine nucleotides (Sandra, John et al. 2000). DMS footprinting has been demonstrated *in vivo* and is capable of detecting secondary and tertiary structure. However, the method is time consuming and takes 1.5-3 days for 300-500 nucleotides. Importantly, DMS is highly toxic, volatile, is a carcinogen and is readily absorbed through the skin making it difficult and dangerous to work with. The biggest disadvantage of DMS though is its sensitivity to the A and C bases only, which means structural information at other nucleotides may go undetected.

Finally, SHAPE depends on the enhanced nucleophilicity of the 2'-hydroxyl group in unconstrained nucleotides compared to base-paired nucleotides (Kevin, Edward et al. 2006). SHAPE is sensitive to all four nucleotides and has been demonstrated *in vivo*. Unfortunately, the technique cannot be applied to DNA. Furthermore, since SHAPE involves the complete determination of the secondary structure, the procedure is extremely long and it takes two days to complete a short RNA sequence of 100-200 nucleotides.

The techniques described above attempt to precisely determine secondary or tertiary structure, which makes them laborious and time consuming. More importantly, these techniques use highly toxic chemicals to probe structure, making them ineffective as *in vivo* structural probes. Hydroxyl radical footprinting modifies the ribose or deoxyribose backbone, DMS is a carcinogen that modifies the nucleotides and SHAPE makes use of electrophiles, which necessitates the extraction of the cell from the human body before the structure analysis can take place. Furthermore, these techniques employ heating, freezing, drying, mixing, centrifugation or sequencing reactions that require that the DNA/RNA of interest be extracted from the cell. Therefore, the above-mentioned techniques cannot be used to detect the presence of a particular structure in the human body nor can they be used for treatment of the cells containing the aberrant structure. However, if there is a priori knowledge that the cell that is malignant and the structure of the DNA/RNA elements is unknown, the appropriate technique mentioned above can be used to determine the secondary or tertiary structure outside the human body.

11.3. Secondary and tertiary structure detection using the RET network

The RET network we designed, self-assembles DNA probes labeled with acceptor fluorophores to the target DNA/RNA secondary or tertiary structure forming an optical network. A large number of studies have shown that foreign DNA can bind to regions in the nucleosome with high specificity. This specificity combined with the small size of fluorophores allows easy access to highly condensed regions of the nucleosome (Yaroslavsky and IV 2013). The number of probe strands required varies depending on the complexity of the structure being detected. A DNA probe strand labeled with a donor fluorophore binds to a unique sequence adjacent to the secondary/tertiary structure. The region adjacent to the target structure can be a single stranded region like a loop, bulge or an overhang but it can also be a duplex region to which the probe strand can triplex bind. Triplex binding involves non-canonical base pairing between double stranded RNA and an additional single-stranded RNA molecule. In (Zhou, et. et al. 2013), the authors successfully demonstrate the binding of a locked nucleic acid, 2-thi U- and 2'-O methyl-modified residues to RNA/DNA and RNA/RNA hairpin structures. Triplex binding has a different melting temperature from that of duplex binding and this temperature is significantly altered with salt concentration. These two properties can be used to control duplex versus triplex formation in vitro. Once the optical network is assembled, the donor fluorophore is excited and the time-resolved fluorescence output

is observed from all the fluorophores in the network. A different time-resolved optical signature is observed based on the presence of the wild-type or the aberrant secondary structure. Such an approach should be able to distinguish between different secondary or tertiary DNA/RNA structures and hence identify certain types of cancer cells with high specificity.

In addition to the advantages of the RET network over conventionally used fluorescence barcodes, the RET network has the following advantages at determining differences in the secondary and tertiary structure:

1. The sensor can differentiate between targets based on their secondary and tertiary structure.

2. The sensor makes use of DNA/RNA probe strands, which makes it very specific to the genome wide location of the target DNA/RNA structure.

3. The use of oligonucleotides and the small size of the fluorophores allow access to highly condensed regions in the nucleosome. The small size of the fluorophore labeled probe strands enables high spatial resolution.

4. Custom oligonucleotides conjugated with fluorophores can be easily purchased. The hybridization protocol for the self-assembly of the probe strands to the target is not labor intensive and not hazardous.

5. The RET network is essentially a nanoscale optical computing system. Therefore, if the fluorescence output indicates that a cell is malignant, the output of the sensor can be used to drive therapeutic agents such as photosensitizers to kill the cell with high specificity.

In this work, we model the RET network to distinguish between normal cells and cells affected with lung and breast cancer, which are the most frequently occurring cancers worldwide in men and women respectively.

11.4. Detecting lung cancer using the RET network

Glutathione peroxidase (GPX3) mRNA is expressed in the kidney, heart, lung, liver, brain, adipose tissue, breast, and gastrointestinal tract and modulates the adverse effects of reactive oxygen species (ROS) in the extracellular environment (Barrett, Ning et al. 2013, Sabarinathan, Wenzel et al. 2014). ROS can damage tissues, DNA nucleotides, suphydryl groups in proteins and cause crosslinking or fragmentation of ribonucleoproteins. Mutations caused by oxidative DNA damage include a wide range of specifically oxidized purines and pyrimidines, single strand breaks and instability formed directly or by repair processes. All four DNA bases can be modified by ROS but the GC base pair is affected more than the AT pair. The damage to DNA by ROS is one of major causes of cancer. Hypermethylation and underexpression, and thus a loss of activation of GPX3, has been associated with lung, prostate, gastric, cervical, thyroid and colon cancers (Barrett, Ning et al. 2013). GPX3 codes for an enzyme, containing selenocysteine, which catalyzes the reduction of hydrogen peroxide and prevents the release of ROS (Sabarinathan, Wenzel et al. 2014). The amino acid that codes for selenocysteine however is the same as the UGA stop codon. The recognition of UGA as the selenocysteine instead of the UGA stop codon is mediated by GPX3's mRNA cis-

acting regulatory element, called the selenocysteine insertion sequence. In lung cancer, it was noticed that the single nucleotide variation, U1552G, located in the selenocysteine insertion sequence alters the secondary structure of the regulatory region and prevents the formation of the enzyme containing selenocysteine. This in turn prevents the reduction of hydrogen peroxide and causes damage to DNA. Figure 70 shows the differences in the secondary structure of tumor suppressor gene, GPX3 between a normal cell and a cell affected with lung cancer. Our goal is to use the RET network to differentiate between the wild-type and cancerous secondary structure, which in turn can help differentiate between a normal cell and a cancerous cell.

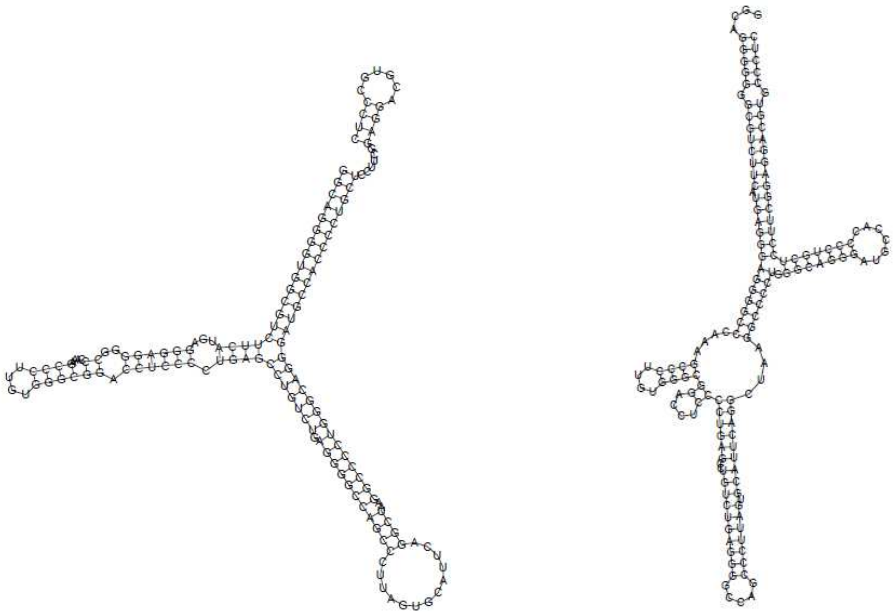


Figure 70: Secondary structure of the regulatory region of GPX3 in wild-type (left U allele) and cancerous cell (right G allele) (Lorenz, Bernhart et al. 2011)

In order to create a RET sensor that can discriminate between the wild-type and cancerous secondary structures, we first create a three-dimensional model of both the wild-type and cancerous strucutures. The RNA sequences and the secondary structure for the wild-type and cancerous strucutures are shown below. The secondary structure is shown in the frequently used dot bracket notation. The notation includes dots, open and closed parenthesis. A dot denotes an unpaired base while an open parenthesis indicates that the base is paired with the base denoted by the matched closed parenthesis.

Wild-type, U allele:

GGCAGGGGUGGCGUCUUCAUGAGGGAGGGGCCCAAAGCCCUUGUGGGCGGAC

CUCCCCUGAGCCUGUCUGAGGGGCCAGCCCUUAGUGCAUUCAGGCUAAGGCC

CCUGGGCAGGGAUGCCACCCCUGCUCCUUCGGAGGACGUGCCCUC

((((((((((((((((((...((((((..((....((((....))))))..)))))).))))((((((.((((((((((.....))))..))))))))))))))))))))

))))......((((.....))))

Lung cancer, G allele:

GGCAGGGGGGGCGUCUUCAUGAGGGAGGGGCCCAAAGCCCUUGUGGGCGGAC

CUCCCCUGAGCCUGUCUGAGGGGCCAGCCCUUAGUGCAUUCAGGCUAAGGCC

CCUGGGCAGGGAUGCCACCCCUGCUCCUUCGGAGGACGUGCCCUC

....(((((.(((((((((.(((((((((((((....((((....))))((....))((((((..((((((((((.....)))))))..)))))))))....)))))))).(((((((.....)))))
)))))))))))))))).)))))

11.4.1. 3-dimensional structure model

RNA has a large conformational landscape due to its six backbone dihedrals per residue. For an RNA of length N, approximately 3^{6N} conformations exist (Leontis 2012). For sequences longer than 30 bases such as the long lung cancer sequences shown above, coarse-grained methods are inaccurate while full-atom molecular dynamics models, which identify the globally optimal structure, require prohibitively large computational resources. In order to address these issues, two alternate modeling techniques have emerged: homology modeling and fragment based assembly (Leontis 2012). Homology modeling uses a 3-dimensional template structure, which has a high level of sequence or secondary structure similarity with the query sequence. Homologous sequences or secondary structures are known to have similar global 3-dimensional structure. Using homology modeling, a homologous template structure is first identified and local changes are subsequently made to the query structure based on any substitutions, insertions and deletions that exist in the query sequence relative to the template sequence. For the lung cancer sequences shown above, we undertook an extensive and comprehensive homology search of sequence and secondary structure through the genomes and PDB databases of all organisms. However, we were unable to find a

template with a high homology level with the query. A known disadvantage of template-based modeling is the inaccuracy of the query structure if a highly homologous template structure is not found (Leontis 2012). We therefore use the fragment based assembly model to construct 3-dimensional models of the wild-type and lung cancer sequences of interest. The fragment assembly model combines physics-based folding with known 3-dimensional structures of fragments of the query sequence. The input secondary structure is first fragmented and 3D structural elements for each of the fragments are identified. The 3-dimensional structural elements are chosen in the following order: secondary structure, topology, sequence similarity, pyrimidine/purine compatibility, source structure resolution and energy (Popenda, Bielecki et al. 2006, Popenda, Szachniuk et al. 2011). The dictionary of 3-dimensional structural elements contains 190,928 elements. In the event that 3D structure for a certain sub-sequence is unavailable, the 3D structure is predicted using the Nucleic Acid Builder (NAB) (Macke and Case 1997). NAB uses rigid-body transformations, distance geometry, energy minimization, molecular dynamics and normal mode analysis to build 3D structural models. NAB has been used to model duplex, triplex and tetraplex DNA, RNA hairpins, ribosomal subunits and recombination sites. A few dozen to a few hundred nucleotides have been modeled with atomic level resolution thus far. On determinining the 3D structural elements of individual fragments, they are superimposed to yield an initial global 3D model. Finally, the structure is refined using energy minimization techniques

such as Cartesian and torsion angle dynamics and gradient-based minimization (Popenda, Bielecki et al. 2006). The fragment assembly technique has been shown to model RNA sequences of up to 500 bps and different structural complexity including first-order pseudoknots, branched RNA's and tRNA's. The average global root mean square deviation of 40 such models, when compared with their high-resolution X-ray equivalents, is 5.1 Å. The fragment assembly technique reduces the high computational cost of sampling all the probable conformations, improves the accuracy of prediction and takes into account environmental factors that affect structure. Screen shots of the 3D models of the wild-type and aberrant GPX3 structure are shown in Figure 71 (Humphrey, Dalke et al. 1996).

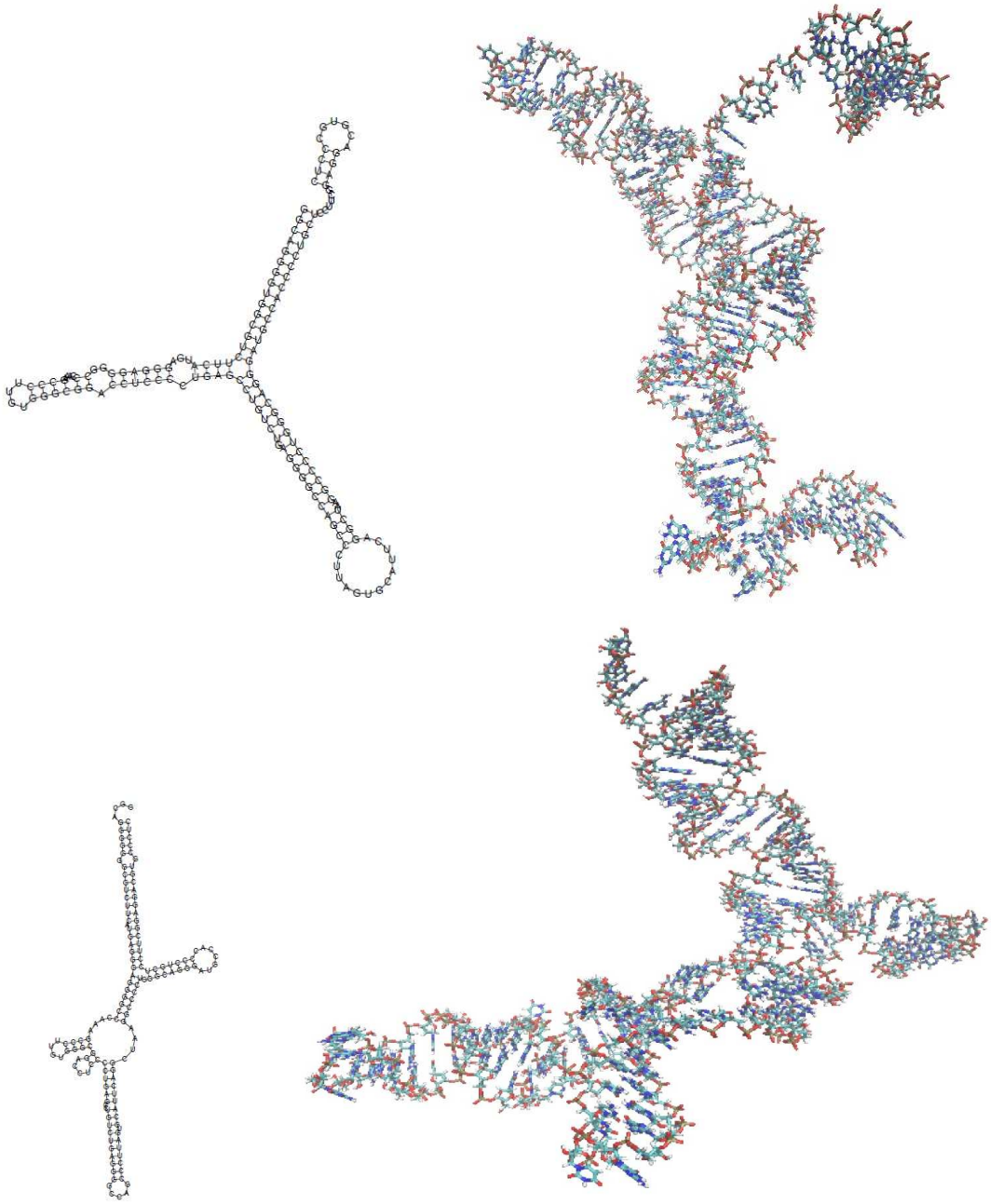


Figure 71: Three-dimensional models of the regulatory region of GPX3 in wild-type (top U allele) and cancerous cell (bottom G allele).

In order to test the robustness of our modeling technique to the initial condition, we modeled the 3D structure for the U allele using the secondary structure for the G allele. We were unable to recover the original 3D structure of the U allele but the structure obtained was not that of the G allele either, indicating that the initial condition is important but does not solely dictate the structure of the backbone. We got a similar result on modeling the 3D structure of the G allele using the secondary structure for the U allele.

11.4.2. Sensor design

The RET sensor is designed to identify parts of the secondary/tertiary structure unique to the cancerous cell compared to the normal cell. Fluorophore conjugated DNA strands hybridize to unique single stranded regions in the cancerous structure. The donor fluorophore (AF 405) and acceptor fluorophores (AF 488, AF 546, AF 555, AF 594, AF 647 and AF 680) form a network around GPX3's structure in the cancerous cell as shown in Figure 72. The fluorophore network is selected such that the most blue fluorophore serves as the donor, which transfers its energy through mediating fluorophores to the most red fluorophore in the network in the following order: AF 405 -> AF 488 -> AF 546 -> AF 555 -> AF 594 -> AF 647 -> AF 680. Energy transfer between each pair of fluorophores is accompanied by a loss of energy and thus an increase in the

emission wavelength of the acceptor flurophore. Therefore, the flurophore at the end of the cascade emits red light, which can be used to drive photo-sensitive therapeutic agents as described in Chapter 12. GPX3's single stranded regions in the aberrant structure are hybridized in its wild-type structure. Therefore, the labeled DNA strands cannot find an accessible single stranded region in the wild-type GPX3 structure. It should be noted here that all the flurophores in the network are positioned to sample crucial changes in global structure in the cancerous cell compared to the normal cell. For detecting local changes in structure, fluorophores can be placed in closer proximity around the structural feature of interest.

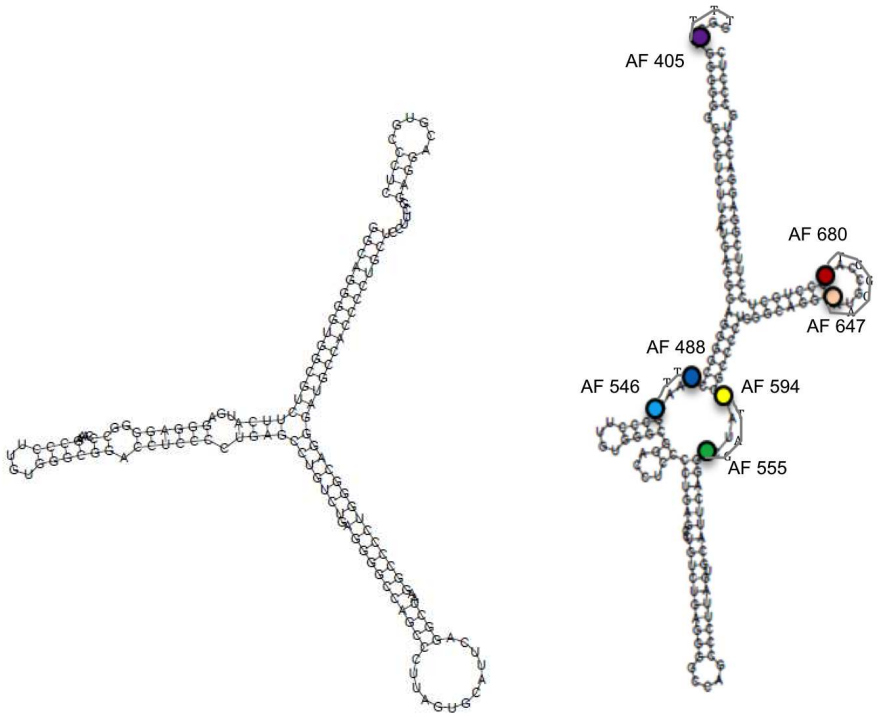


Figure 72: Sensor design for differentiating between wild-type (left) and lung cancer cells (right). The fluorophore indicated with the blue dot is AF 488, the fluorophore indicated with the green dot is AF 594 and the red dot indicates AF 647

To detect the aberrant GPX3 structure, it is hybridized using the following acceptor conjugated sequences:

AF 546/TTTG/AF 488

AF594/TTAG/AF 555

AF 680/GTGGCAT/AF 647

The donor fluorophore is conjugated to a region adjacent to the secondary/tertiary structure. In theory, a fragment of DNA over 16 bases long should be a unique sequence

in the genome (Cantor and Cassandra 2004). However, there exist DNA sequences that are over 16 bases long and yet appear in multiple regions within the genome. Using a long sequence increases the probabilty of the sequence being unique. However, a long sequence also increases the probability of subsequences (n>5) binding to regions of the genome that does not contain the target structure. In order to accommodate this trade off, we chose the following sequence adjacent to the 5′ end of the target sequence: GCAAGGGCCACGGACCCCATGGCA. This sequence does not exist in any other location in the genome even at 95% similarity. However, several subsequences of the sequence exist in different regions of the genome. The DNA strands conjugated with the acceptor fluorophore are not present within R_o of the donor fluorophore on either the 5′ or the 3′ end of the subsequences, which results in minimal excitation of the acceptor fluorophores bound to regions other than the target strucure. Therefore, when the donor fluorophore is excited at 405 nm, a large part of the energy transfer is to the acceptor fluorophores in the regulatory region of GPX3. Furthermore, none of the acceptor fluorophores are excited directly at the excitation wavelength of the donor fluorophore. Therefore sub-networks formed due to the aberrant binding of acceptor labeled fluorophores to regions of the genome that do not contain the target structure will not affect the output fluorescence signature.

We simulate the outputs of the RET network using the Markov model described in Chapter 3. Using the PDB files of the 3D models, we are able to generate coordinates for every atom in the wild-type and aberrant structure. The fluorophores we typically use have an amino modification on the 5′ end of the sequence from the P position of the phosphate. Therefore, the coordinates of the P position of the base to which each fluorophore is conjugated is taken as the position of the 2 nm linker and flurophore. The sensor is excited at 405 nm and the fluorescence output from all the fluorophores is observed. In the case of the aberrant structure, the external probe sequences will bind to the single stranded regions of the target structure bringing the acceptor flurophores in close proximity to the donor flurophore, AF 405. This should result in a significant transfer of AF 405's energy to the acceptor fluorophores since energy transfer is inversely proportional to the sixth power of the distance between the donor and acceptor fluorophores. In the wild-type structure, the single stranded RNA sequences in the aberrant structure are hybridized and are unavailable for binding by the external DNA probe strands. In this case, the optical output will correspond to the fluorescence from AF 405 with no acceptor fluorophores in the vicinity, as indicated by the red histogram in Figures 73 - 79. It is clear that the optical output corresponding to the wild-type and aberrant secondary structure is significantly different. As such, independent of the fluorophore being observed, the normal and cancerous cells are easily distinguishable.

Table 10: PDB coordinates of the bases to which fluorophores are conjugated.

Base	A	C	A	C	A	A	C
Position	4	33	36	97	100	116	122
X coord.	15.052	81.913	77.050	84.731	73.579	72.135	74.116
Y coord.	-3.408	11.678	25.178	-0.900	-1.178	-23.658	-32.374
Z coord.	4.477	-19.876	-10.818	5.559	-6.416	-43.677	-30.205

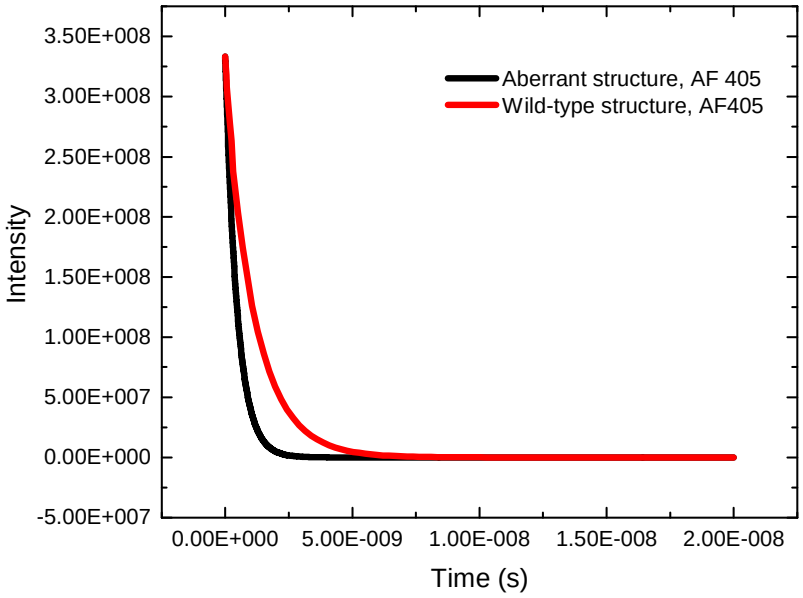


Figure 73: Time resolved fluorescence histograms from the donor fluorophore, AF 405, corresponding to the wild-type and aberrant secondary structure.

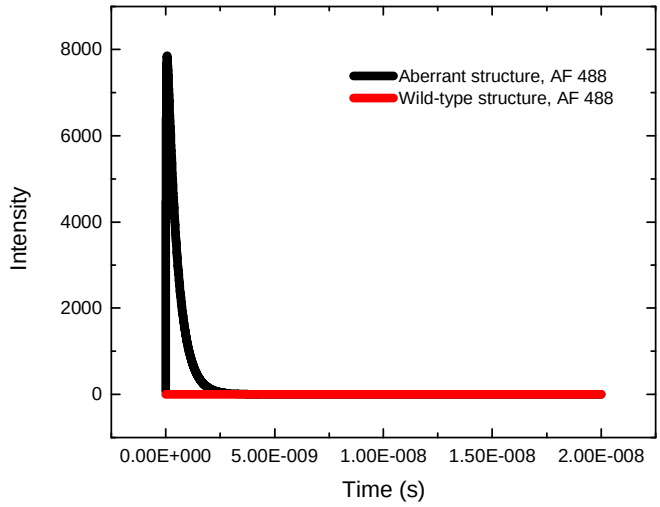


Figure 74: Time resolved fluorescence histograms from the acceptor fluorophore, AF 488, corresponding to the wild-type and aberrant secondary structure.

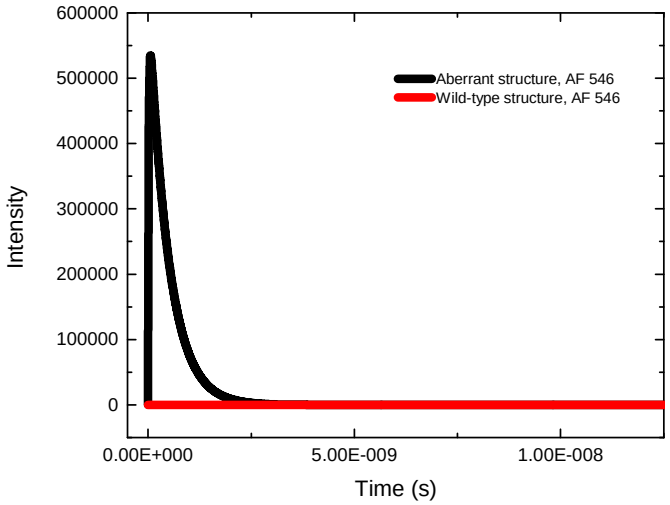


Figure 75: Time resolved fluorescence histograms from the acceptor fluorophore, AF 546, corresponding to the wild-type and aberrant secondary structure.

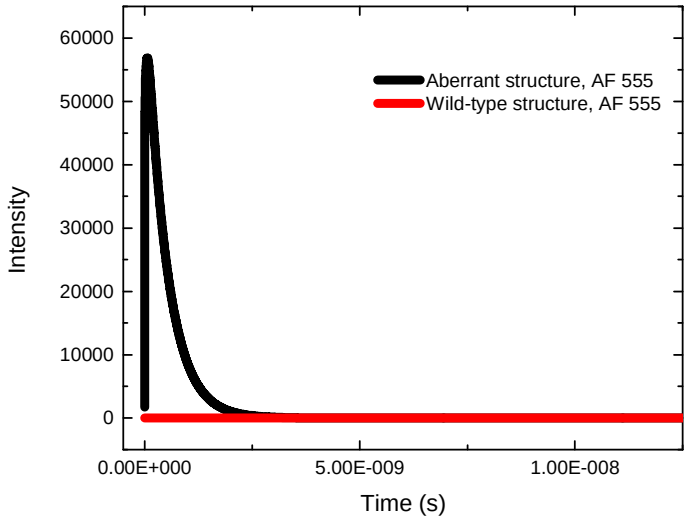


Figure 76: Time resolved fluorescence histograms from the acceptor fluorophore, AF 555, corresponding to the wild-type and aberrant secondary structure.

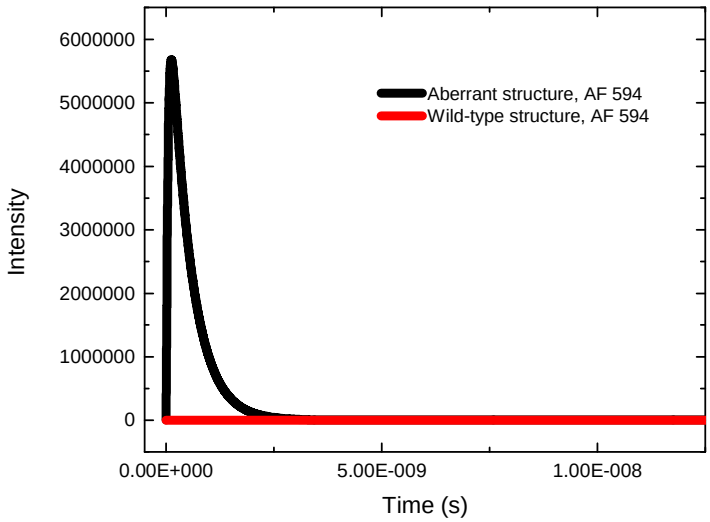


Figure 77: Time resolved fluorescence histograms from the acceptor fluorophore, AF 594, corresponding to the wild-type and aberrant secondary structure.

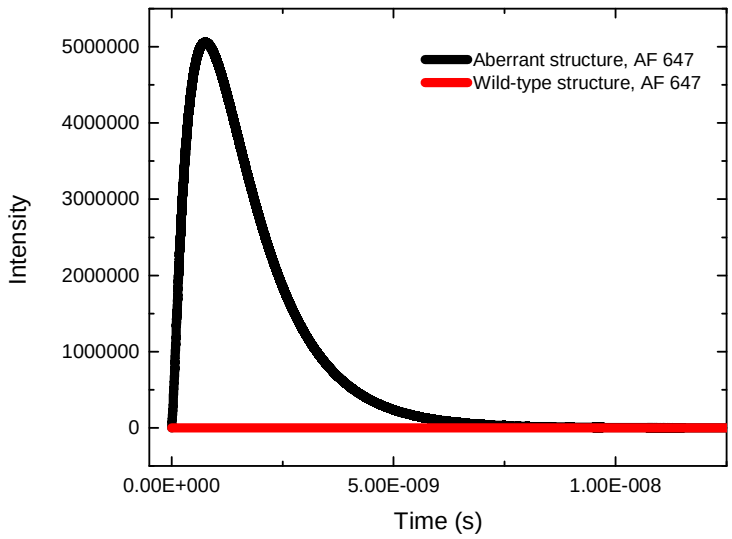


Figure 78: Time resolved fluorescence histograms from the acceptor fluorophore, AF 647, corresponding to the wild-type and aberrant secondary structure.

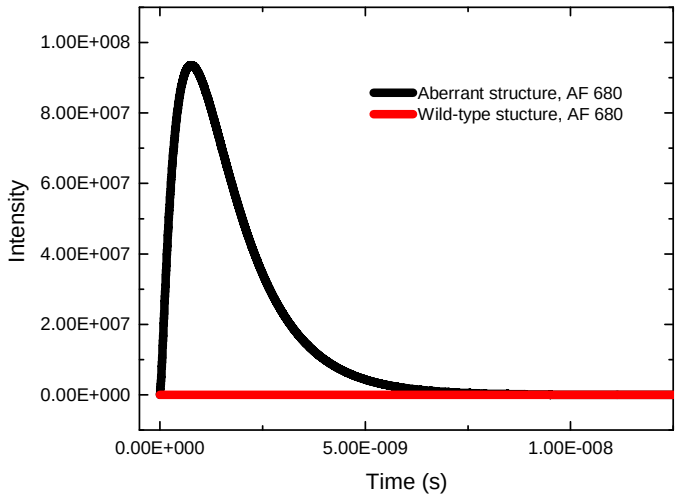


Figure 79: Time resolved fluorescence histograms from the acceptor fluorophore, AF 680, corresponding to the wild-type and aberrant secondary structure.

11.5. Detecting breast cancer using the RET network

BRCA2 is a tumor suppressor gene involved in cell proliferation, development, DNA damage repair, transcription and centrosome duplication (Sailesh, Bukhari. et al. 2007). Defects in the BRCA gene activate the p53 pathway and its downstream target p21, which in turn halts the cell cycle progression. Loss of function of this tumor suppressor gene or abnormal increase in gene expression is often implicated in breast cancer (King, Marks et al. 2003). Due to the large number of processes BRCA2 is involved in, its expression is very tightly regulated. Polymorphisms in the regulatory region of BRCA2 can disrupt the fine-tuned regulation of its expression. A polymorphism in the 5′ untranslated region of the BRCA2 gene has been implicated in breast cancer (Sailesh, Bukhari. et al. 2007). The G allele has a low expression level and inhibits DNA repair, which can lead to genomic instability. The A allele, on the other hand, is found to increase the expression levels, which leads to inhibition of p53 transactivation and can contribute to high genomic instability. Therefore, intermediate levels of BRCA2 expression, facilitated by heterozygous A and G alleles are essential to maintain normal cell response to DNA damage. A change in the polymorphism from G to A at the site -26 was found to double the gene expression in MCF-7 and HeLa cells (Sailesh, Bukhari. et al. 2007). It is predicted that the altered stability of the secondary structure in the vicinity of the translation start site and increased translation efficiency is

the cause for the A allele resulting in breast cancer. The RNA sequences and the secondary structure for the wild-type and cancerous strucutures are shown below:

Wild-type, A allele:

ACCUCUGGAGCGGACUUAUUUACCAAACAUUGGAGGAAUA

.(((((((...((.........)).....)))))))....

Cancerous, G allele:

ACCUCUGGAGCGGACUUAUUUACCAAGCAUUGGAGGAAUA

.(((((((.((.............)).)))))))....

In order to create a RET network that can discriminate between the wild-type and cancerous secondary structures, we first created a 3D model of both the wild-type and cancerous structures. The 3D models are created using the fragment based assembly

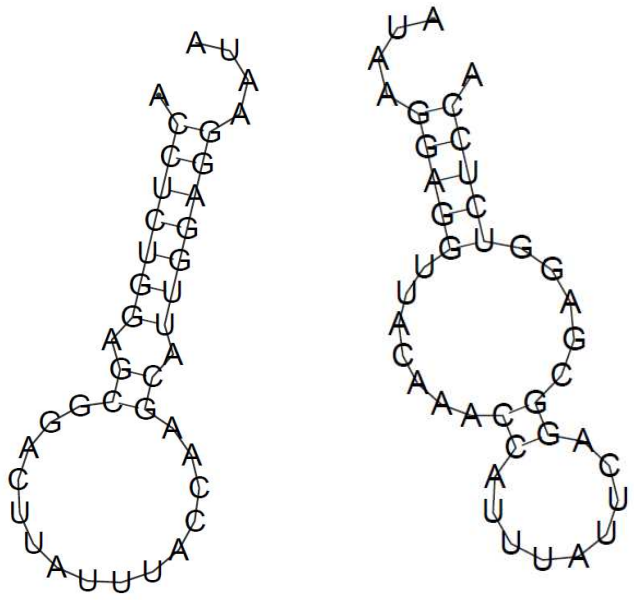


Figure 80: BRCA2, secondary structure variation in the wild-type A allele (left) and the cancerous G allele (right).

approach described earlier. Again, in order to test the robustness of our modeling technique to the initial condition, we modeled the 3D structure for the A allele using the secondary structure for the G allele. We were able to recover the 3D structure of the A allele despite using the initial condition for the G allele. On modeling the G allele with the secondary structure for the A allele, we did not recover the 3D structure for the G allele but the structure obtained was not that of the A allele either, indicating that the initial condition is important but does not solely dictate structure. The 3D models of the wild-type and cancerous structures are shown in Figure 81 (Humphrey, Dalke et al. 1996).

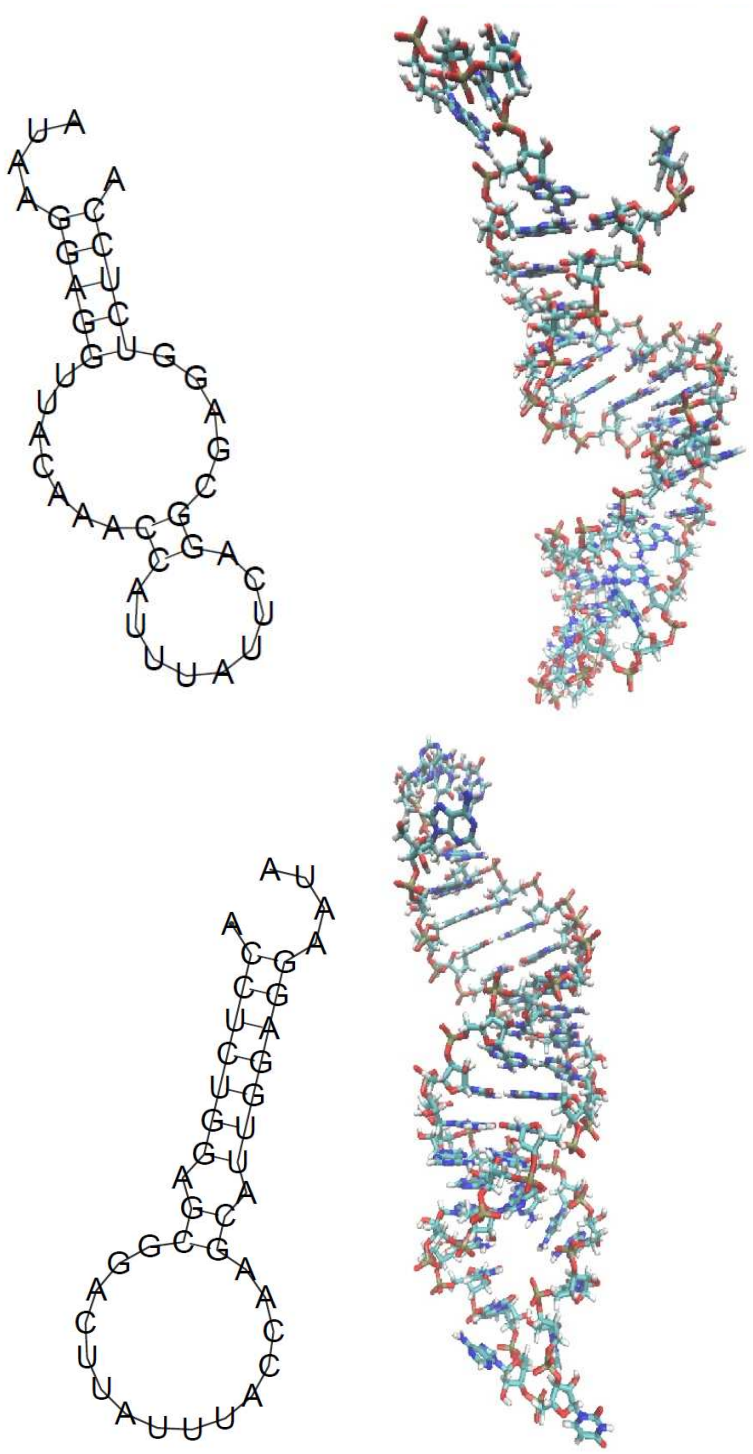


Figure 81: Secondary structure in the untranslated region of BRCA2 in wild-type G allele and cancerous A allele.

11.5.1. Sensor design

Similar to the earlier lung cancer design, the RET network is designed to identify parts of the secondary structure unique to the breast cancer cell compared to the normal cell. Fluorophore conjugated DNA strands hybridize to unique single stranded regions in the aberrant BRCA2 structure. The donor fluorophore (AF 405) and acceptor fluorophores (AF 488, AF 546, AF 555, AF 594, AF 647 and AF 680) form a network around the structure of BRCA2 in the cancerous cell as shown in the Figure 82. The flurophore network is selected such that the most blue fluorophore serves as the donor flurophore and transfers its energy to mediating flurophores placed at across the aberrant structure. The mediating fluorophores are positioned such that neighboring fluorophores are good donor-acceptor pairs. When the donor fluorophore is excited with blue light, energy is lost at each energy transfer stage accompanied by an increase in the wavlength of the emission spectra of the acceptor fluorophore. Therefore, the fluorophore at the end of the RET cascade emits red light, which can be used to drive therapeutic agents as described in Chapter 12. Two out of the three single stranded regions in the cancerous cell are hybridized in the normal cell. Therefore, even when the same set of labeled probe strands are used, the as-formed RET network differs in the wild-type and aberrant structure.

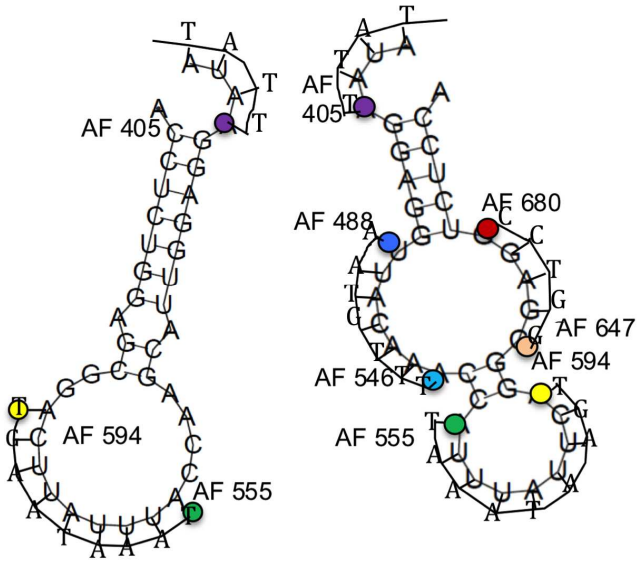


Figure 82: Sensor design for differentiating between wild-type (left) and breast cancer (right) cells.

To detect the aberrant BRCA2 structure, the structure is hybridized using the follwing acceptor conjugated sequences:

AF 680 GCTCC AF 647

AF 594 TAAATAAGT AF 555

AF 546 AATGTTT AF 488

The donor fluorophore is conjugated to the following unique region adjacent to the secondary structure: AATATCGTAGGTAAAAATGC. This exact sequence does not exist in any other location in the genome though several subsequences exist. However, the DNA strands conjugated with the acceptor fluorophore are not present within $2R_o$ of

the donor fluorophore on either the 5′ or the 3′ end of the subsequences, which ensures that there is negligable excitation from the acceptor fluorophores that non-specifically attach to different regions of the genome. Therefore, when we excite the donor fluorophore at 405 nm, it can only excite the acceptor fluorophores in regulatory region of BRCA2. Furthermore, exciting the donor fluorophore at a wavelength of 405 nm ensures that none of the acceptor fluorophores are directly excited without the presence of the donor in the vicinity. This ensures that the optical output is not affected due to sub networks formed by the aberrant binding of acceptor labeled fluorophores to regions of the genome that do not contain the target structure.

We simulate the outputs of the RET network using the Markov model described in Chapter 3. Using the PDB files of the 3D models, we are able to generate the coordinates for every atom in the wild-type and aberrant structures as shown in Table 11 and Table 12. The coordinates of the P position of the base to which each fluorophore is conjugated is taken as the position of the 2 nm linker and flurophore. The sensor is excited at 405 nm and the fluorescence outputs from all the fluorophores are observed. In the case of the aberrant structure, the external probe sequences will bind to the single stranded regions of the target structure bringing the acceptor flurophores in close proximity to the donor flurophore, AF 405. This should result in a significant transfer of AF 405's energ to the acceptor fluorophores. The acceptor fluorophores in turn emit a

high intentisty fluorescence signal as shown in Figures 83 - 89. In the wild-type BRCA2 structure, two out of the three labeled probe strands do not have an accessible binding site. As such, the fluorophores AF 488, AF 546, AF 647 and AF 680 cannot hybridize to

Table 11: PDB coordinates of the positions where fluorophores are attached in the A allele of BRCA2.

Base	A	G	C	A	A	A	U	A	A
Position	1	7	11	14	22	25	31	37	40
X coord.	0.0881	2.785	-19.58	-12.551	-6.698	-1.059	-10.243	15.386	23.344
Y coord.	1.724	-12.586	-15.62	-31.382	-21.328	-22.868	-2.339	0.675	7.710
Z coord.	-0.544	25.824	26.376	28.683	44.063	26.994	16.930	9.801	5.183

Table 12: PDB coordinates of the positions where fluorophores are attached in the G allele of BRCA2.

Base	A	A	A
Position	14	22	37
X coord.	-12.551	-6.698	15.386
Y coord.	-31.382	-21.328	0.675
Z coord.	28.683	44.063	9.801

the wild-type structure and do not emit any fluorescence. The only remaining probe strand, AF 594TAAATAAGTAF 555, can bind to the wild-type structure since its

binding site is still single stranded. The R_0 of AF 405 and AF 555 is 3.6 nm while that of AF 405 and AF 594 is 2.6 nm. The distance between AF 405 and AF 555, AF 594 is beyond that of their R_0 value, at 4.5 nm and 5.07 nm respectively. Because of the large distance between the donor and acceptor fluorophores relative to their R_0 value, without the mediating donor-acceptor pairs, the fluorescence from AF 555 and AF 594 in the wild-type structure will be significantly lower than the fluorescence obtained from the aberrant structure, as seen in Figure 86 and Figure 87. We see that the time resolved fluorescence output from any of the seven fluorophores in the network can help us discrimimate between the cancerous and normal cells.

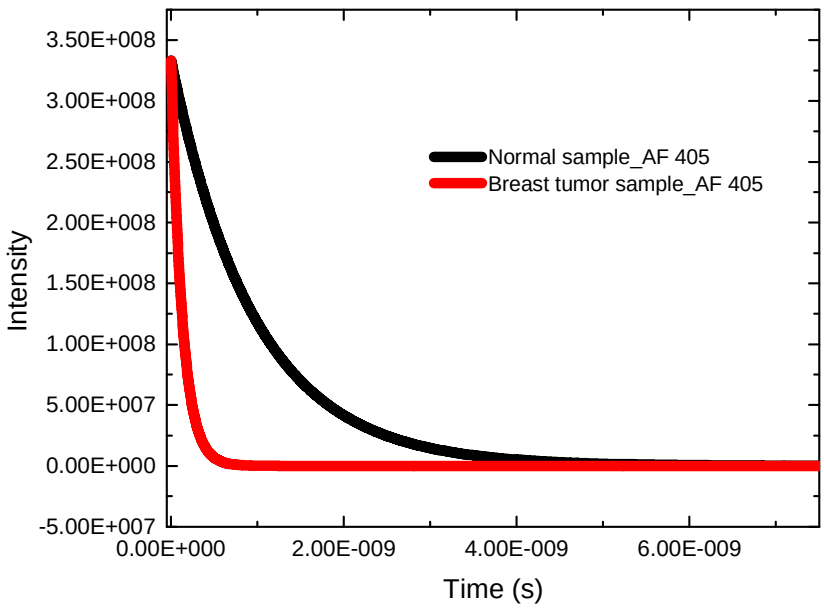


Figure 83: Time resolved fluorescence histograms from the donor fluorophore, AF 405, corresponding to the wild-type and aberrant secondary structure.

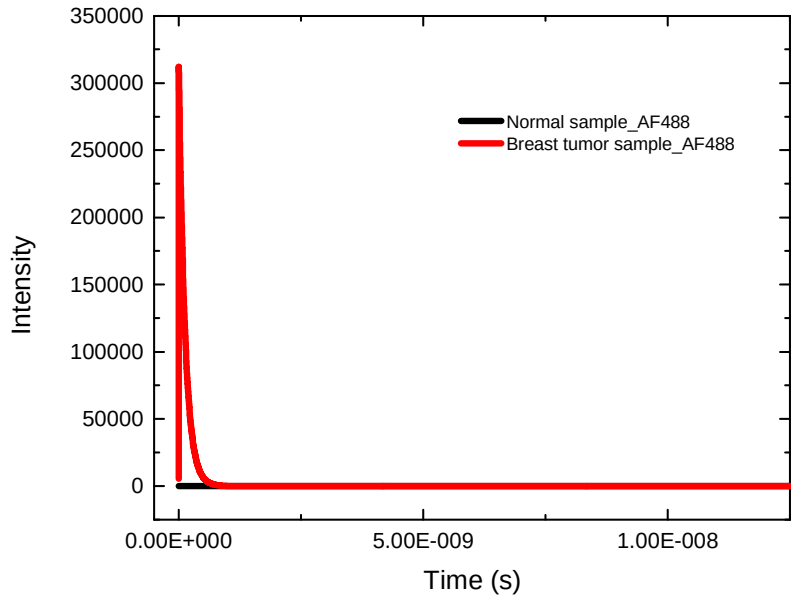


Figure 84: Time resolved fluorescence histograms from the acceptor fluorophore, AF 488, corresponding to the wild-type and aberrant secondary structure.

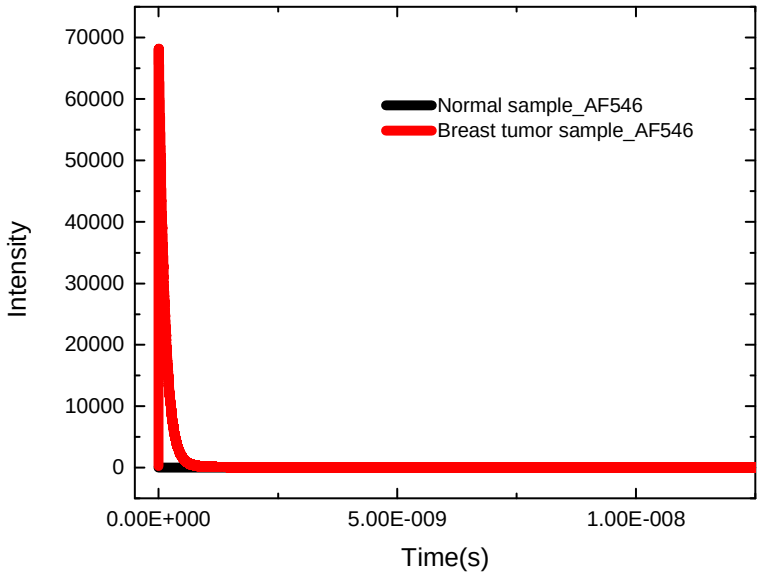


Figure 85: Time resolved fluorescence histograms from the acceptor fluorophore, AF 546, corresponding to the wild-type and aberrant secondary structure.

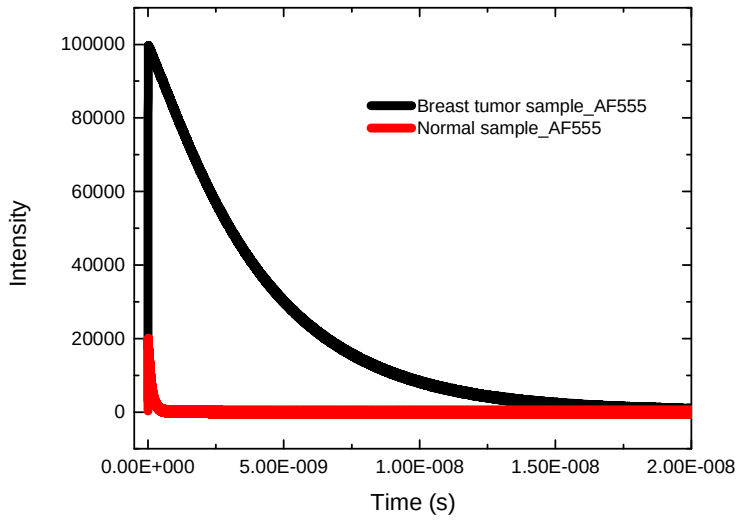


Figure 86: Time resolved fluorescence histograms from the acceptor fluorophore, AF 555, corresponding to the wild-type and aberrant secondary structure.

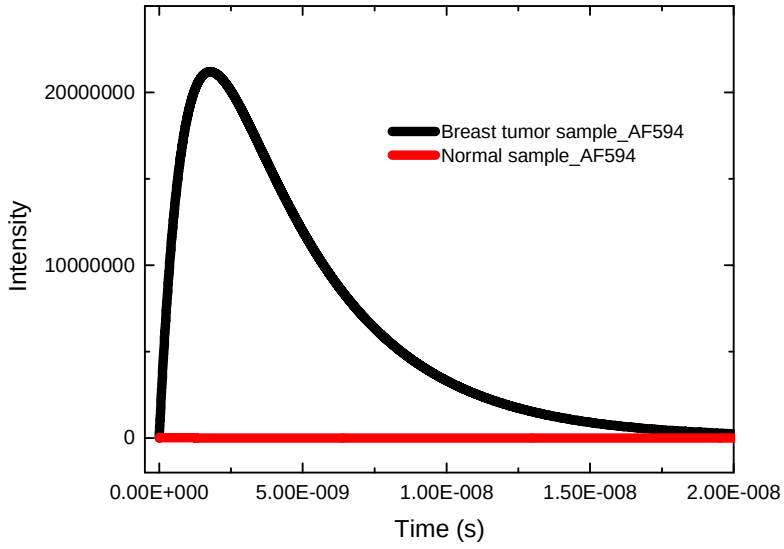


Figure 87: Time resolved fluorescence histograms from the acceptor fluorophore, AF 594, corresponding to the wild-type and aberrant secondary structure.

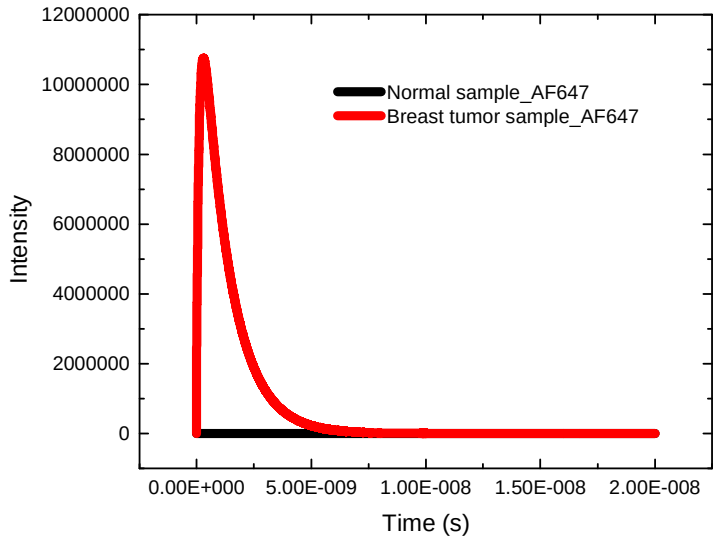


Figure 88: Time resolved fluorescence histograms from the acceptor fluorophore, AF 647, corresponding to the wild-type and aberrant secondary structure.

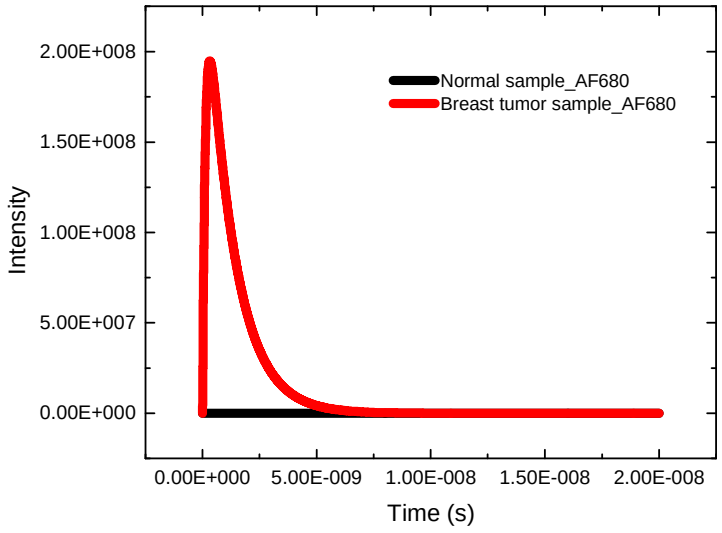


Figure 89: Time resolved fluorescence histograms from the acceptor fluorophore, AF 680, corresponding to the wild-type and aberrant secondary structure.

In both lung and breast cancer detection using the RET network, at least seven different time-resolved signatures are available to differentiate between the wild-type and aberrant secondary/tertiary structure. This results in less than 0.1% false positives and false negatives when the RET network is used to differentiate between cancerous and normal cells, using the analysis techniques described in Chapter 8. It is important to note that it would be extremely difficult to simultaneously characterize, either spectroscopically or using geometrically encoded barcodes, the seven fluorophores used in the above RET networks. The excitation and emission spectra of the seven fluorophores used in lung and breast cancer detection are shown in Figure 90 and Figure 91. There is substantial overlap in the emission spectra of fluorophores, which makes it infeasible to observe the output of any single fluorophore. Furthermore, if the seven fluorophores were to be placed on a geometrical barcode, the barcode would need to be at least 200 nm long and molecular scale spatial resolution cannot be achieved. Therefore, the RET networks, with their small size, high spatial resolution and high information density, are ideally suited to differentiate between the secondary or tertiary structure of the cancerous and normal cells.

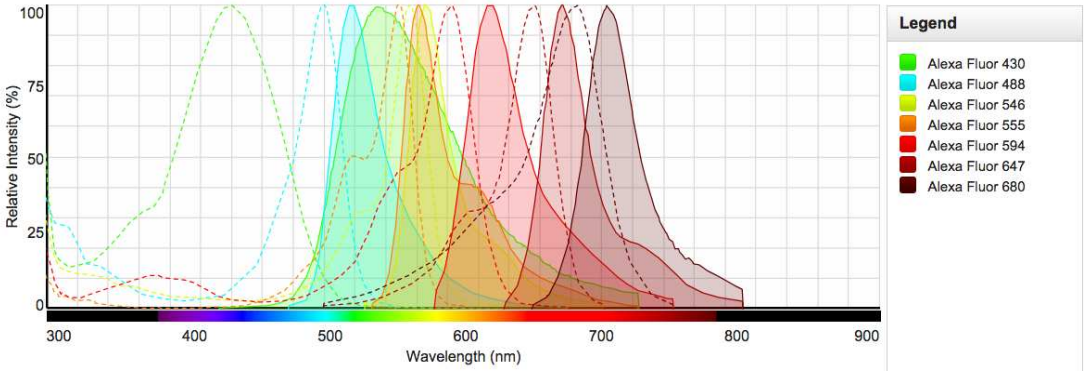


Figure 90: Figure showing spectral overlap between the seven fluorophores used in the RET network for lung cancer detection. Due to the high overlap between the emission spectra of the different fluorophores, it would be very difficult to distinguish between the output signatures of individual fluorophores.

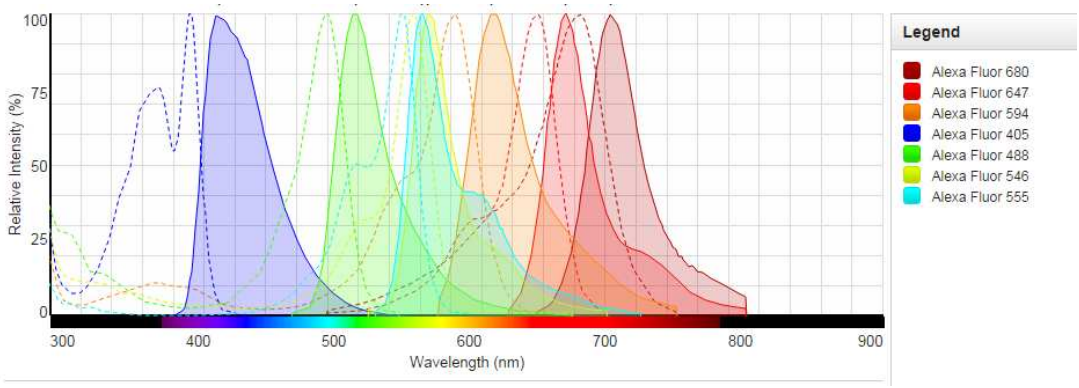


Figure 91: Figure showing spectral overlap between the seven fluorophores used in the RET network for breast cancer detection. Due to the high overlap between the emission spectra of the different fluorophores, it would be very difficult to distinguish between the output signatures of individual fluorophores.

There are also many advantages to using a fluorophore network instead of the conventionally used donor-acceptor fluorophore pair:

1. **Detect local changes in structure**: If there are local changes in secondary or tertiary structure between the wild-type and aberrant structures, multiple donor-acceptor pairs around the local structural change are needed to minimize the number of false negatives. The time-resolved optical output from multiple donor-acceptor pairs can capture the local change in structure with higher sensitivity than a single donor-acceptor pair.
2. **Detect large structures**: The donor fluorophore is conjugated to a unique sequence adjacent to the structure of interest. The secondary or tertiary structure to be probed may not always be close to the donor fluorophore for FRET to occur (typically <10 nm). In that case it helps to have the donor fluorophore transfer its energy to intermediate fluorophores, which in turn can transfer their energy to parts of the distant secondary or tertiary structure of interest.
3. **Minimize false positives**: Having multiple DNA strands labeled with donor and acceptor fluorophores minimizes the possibility of an identical RET network being created at a structure other than the target structure. A single donor acceptor pair greatly increases the probability of non-specific binding, which in turn increases the number of false-positives from the sensor.

In this chapter, we demonstrated, through modeling, that the RET network can be designed such that the optical output from the aberrant structures is high but the wild-

type structure shows negligible fluorescence output in both the lung and breast cancer cases. The optical output from the aberrant structure can therefore be used to drive therapeutic agents, such as photosensitizers, that can selectively kill the cancerous cells with the aberrant structure but leave the normal cells unharmed. The feasibility of using the RET network for cancer therapy and an experimental plan to test the sensor *in vivo* is provided in Chapter 12.

12. Future Work

12.1. Unclonable Cryptographic Key for Secure Authentication

We have demonstrated, through both modeling and an extensive and comprehensive set of experiments, that secure authentication using the RET-key is possible. We created the RET-keys, built a measurement set-up, created a data analysis tool and designed a protocol that will enable authentication between legitimate users of the key but an authentication attempt by an attacker will be unsuccessful. The security system is therefore currently ready for use. However, there are at least a couple of opportunities to increase the usability of the security system and increase its cost-effectiveness. The first is to make a compact, automatic, high speed and accurate detection system that can be integrated with the data analysis tools. TREX currently requires a large bench footprint, is of relatively high cost and requires a manual alignment process. On the excitation side, the entire TREX set-up can be scaled down using the photonic integrated chip technology (Coldren, Corzine et al. 2012). All the individual components including the laser, power splitters, filters, optical amplifiers, optical modulators and detectors are commercially manufactured. On the output side, we already created a digital, high-speed, parallel detection system to detect multiple SPAD signals simultaneously in Chapter 6. This output can easily be integrated with the

data analysis tools described in Chapter 4. The thermal stability of the DNA grids can also be improved so that the measurement environment need not be tightly controlled. The DNA grids are stable up to 37 °C, after which the RET network is lost due to thermal denaturation of the DNA nanostructure. Techniques to improve the thermal stability of the DNA grids can include enzymatic ligation (Erie, Sinha et al. 1987), PNA (Giesen, Kleider et al. 1998), LNA and photocrosslinking (Shigetaka and Kenzo 2014).

12.2. Multiplexed fluorescence sensor for cancer detection

We demonstrated the ability of the RET network to detect molecular changes in nucleotide structure, which was used to distinguish lung and breast cancer cells from normal cells. The ultimate goal of the project, however, is to test the efficacy of the RET network at specifically identifying and treating lung and breast cancers in humans. Towards this goal, we first provide a broad description of how the RET network can be used for theranostic applications and then provide specific details regarding the *in vitro* and *in vivo* experimental procedures.

Testing the RET sensor in humans involves the transportation of the probe DNA strands to the lung and breast tumor cells, the controlled release of the DNA strands once they are inside the cell, testing the specificity and binding efficiency of the probe DNA strands to the target structure and testing the efficacy of the RET network at

driving therapeutic agents. Single stranded DNA conjugated with fluorophores can be injected into a human using nanoparticles with a surface functionalization that ensures high specificity towards the tumor cell of interest. Several different nanoparticles have been formulated to carry single stranded DNA and RNA into tumor cells. These include lipid, polymer and carbohydrate based nanoparticles as well nanoparticles created from inorganic nanomaterials (Mousa and Bharali 2011). The disadvantage of using lipid, polymer and carbohydrate based nanoparticles is their tendency to non-specifically accumulate in the liver and the spleen. Increased specificity can be obtained by using nanoparticles engineered using DNA self-assembly including nanoscale tetrahedrons (Lee 2012), icosahedrons (Bhatia 2011) and octahedrons (Shih, Quispe et al. 2004). These nanoparticles are endocytosed into tumor cells during interphase and release labeled probe ssDNA into the cell. Multiple triggers have been demonstrated to release DNA from the nanoparticles including pH changes, temperature changes, conformational switching and chemical or photon-induced triggers (Mura, Nicolas et al. 2013). Once the labeled probe strands enter the cell, their design ensures that they specifically self-assemble to the aberrant nanostructure and form a nanoscale optical network. In order to activate the optical network, subcutaneous cancers are directly excited with a laser at the excitation wavelength of the donor fluorophore while internal cancers are excited using an endoscope (Wang and Dam 2004), optical fiber (Caspers, Lucassen et al. 2003) or bioluminescence excitation (Zhou, et. et al. 2013). On exciting the donor fluorophore,

energy is transferred to the acceptor fluorophores through RET. The design of the probe sequences and the molecular size of the target structure ensure that acceptor fluorophores in the cancerous cell release a high intensity optical signal. The optical signal released from the cancerous cell is of the same wavelength as that needed to drive a photosensitizer, which releases oxygen free radicals and kills the cancerous cells. The normal cells do not have the same secondary/tertiary structure as the cancerous cells, which results in the fluorophore conjugated probe strands forming a different optical network in the normal cell compared to the cancerous cell. The design of the probe strands therefore, ensures that the optical signal released from the wild-type network is of insufficient intensity and has a wavelength different from that needed to excite the photosensitizer. As such, the normal cells remain unharmed.

In the process described above, the contribution of this thesis is in the creation of the self-assembled nanostructure and the use of the RET network to detect molecular changes in a DNA/RNA structure in a cancer cell relative to a normal cell. In order to test the sensor *in vivo*, a wide range of experimental procedures need to be carried out. Fortunately, most of these experimental procedures have been successfully demonstrated in other systems.

As a first step, the native structure of the wild-type and mutant alleles in lung and breast cancer need to be characterized *in vitro* and *in vivo*. In vivo characterization is required since the RET network is designed to bind to the wild-type and aberrant

structure in live cells. As such, it is important to verify that the aberrant structure exists in the cell's physiological environment. *In vitro* characterization makes it easy to optimize the sensor for binding efficiency and to test the efficacy of the RET network along with the photosensitizer for therapy. There have been reports of identical sequences demonstrating structural differences *in vitro* and *in vivo*, which necessitates both *in vitro* and *in vivo* structure characterization studies (Rouskin 2013). Structure characterization can be performed using RNA SHAPE since both the lung and breast cancer structural aberrations are in the RNA sequence. Secondary structure characterization should suffice to ascertain whether the cancer related alleles result in a different structure from the allele in the normal cell. On verifying that the cancer related alleles indeed alter structure *in vivo* and *in vitro*, the RET network can be tested *in vitro*. For the *in vitro* characterization, the aberrant and wild-type structure are first self-assembled and labeled probe DNA strands are introduced into the system to test for three occurrences: 1. Ensure that the single stranded regions of the aberrant structure are accessible to the labeled ssDNA. 2. Ensure that the binding of the probe strands to the target structure does not disrupt the original structure. 3. Ensure that the labeled probe strands remain bound to the target structure. All three of these tests can be accomplished using FRET donor-acceptor pairs. The design for the FRET pairs is the same as the design to differentiate between the wild-type and cancerous structure with two differences: 1. For the *in vitro* measurements, it is advisable to test each labeled

strand individually while monitoring the fluorescence from the remaining donor acceptor pairs. 2. The donor acceptor pairs, except for the probe strand being tested, should be preconjugated to the original target strand. This will ensure that the stability of the target structure at all vulnerable junctions is tested due to the addition of each individual strand. If the detected levels of fluorescence from all the donor-acceptor pairs follow Försters rule, we can assume that the probe strands are able to hybridize to the target strand and that the original structure is not disrupted. Next, fluorescence from each of the fluorophores at various time points should be observed to test the stability of the structure. Once we verify that the RET network can differentiate between wild-type and aberrant structure, photosensitizers can be introduced into the network. Photosensitizers are molecules, which when activated by a specific wavelength of light release oxygen free radicals. As mentioned earlier, oxygen free radicals can result in preoxidative reactions that can result in cell death. Photodynamic therapy using the photosensitizing agent porfimer sodium has been administered in humans and received FDA approval for the treatment of certain cancers (Usuda, Kato et al. 2006). Porfimer sodium is excited at 630 nm for treatment and therefore can be used at the output of the RET network formed in the lung and breast cancer cells. Techniques to bind porfimer sodium specifically to DNA need to be investigated. The purpose of introducing the photosensitizer at the *in vitro* characterization stage is to test whether the output from the RET network is of sufficient intensity to activate the photosensitizer. The

fluorescence output obtained from the RET network could be significantly higher than the fluorescence that reaches the photosensitizer due to light scattering. The selection of the photosensitizer and its placement with respect to the fluorophores at the end of the RET cascade are both crucial in activating the photosensitizer and can be chosen appropriately during the *in vitro* characterization stage. These choices will need to be further optimized during *in vivo* characterization. *In vivo* characterization requires the injection and controlled release of labeled ssDNA into the tumor cells with high specificity. Nanoparticles containing ssDNA have been observed to circulate through the blood stream and accumulate with high specificity in tumor cells when injected into the tail vein of nude mice (Lee 2012). Nanoparticles such as tetrahedrons and icosahedrons were also able to encapsulate cargo and it was observed that the functionality of the cargo is preserved post delivery, *in vivo*. Other studies have inserted DNA into a phagemid vector, transformed it into a bacterium, which was then inserted into mammalian cells (Lin, Rinker et al. 2008). Once the ssDNA are released inside the cell, it has been previously demonstrated in (Yaroslavsky and IV 2013) that DNA sequences are extremely efficient at accessing highly condensed regions of the nucleosome and are highly specific towards their complementary sequences. This work also showed the assembly of multiple fluorophore labeled ssDNA onto the same target structure. DNA and RNA nanostructures have also been assembled *in vivo*. Single stranded DNA amplified in the bacteria cells have been assembled into a 4-arm cloverleaf structure

using helper phages (Lin, Rinker et al. 2008). Once the ssDNA are released into the cell, it is possible for DNAses to cut the probe DNA strands prematurely before the strands reach their complimentary regions in the target structure. This can be avoided by incorporating LNA into the DNA probe strands (Freiden, Hansen et al. 2003, Sorensen, Nielsen et al. 2004, Valoczi, Hornyik et al. 2004). Nanoscale structures, such as the tetrahedrons and icosahedrons, have been shown to be stable in mice for a few hours before they are degraded. Several hours time is sufficient to probe the assembled RET network since only a few minutes are required to take time correlated fluorescence measurements. If the probe strands are able to find their target structure and form the RET network successfully, the structure can be excited using the excitation wavelength of the donor fluorophore. For subcutaneous cancers, it is easy to target a specific wavelength of light to the tumor site. For internal cancers, light is often directed with the help of small fiber-optic cables, endoscopes or bioluminescence sources to the specific site where the RET networks have accumulated. Whole body imaging can be performed to observe the fluorescence from all the fluorophores in the RET network and identify the tumor (Piper, Habermehl et al. 2013). However, the amount of fluorescence needed for whole body imaging is currently very large. Fortunately, for every gene, multiple RNA transcripts exist per cell and multiple cells with the aberrant structure are clustered together in a tumor. This should greatly increase the probability of detecting fluorescence from the tumor. For therapeutic applications, the fluorescence required

from the RET network is significantly lower than that required for imaging and is determined by the number of photons required to activate the photosensitizer. During *in vivo* characterization if it is determined that the fluorescence output from the RET network is insufficient to drive the photosensitizer, the RET network can be redesigned to have multiple fluorophores at the end of the RET cascade excite the photosensitizer simultaneously.

A comprehensive set of *in vitro* and *in vivo* experiments should therefore provide us with information regarding the ideal design of the RET network at detecting a specific structure, the experimental procedures most suited for the delivery and self-assembly of the labeled DNA probes and the optimum experimental parameters that can improve the specificity, stability and efficacy of the RET network at identifying and selectively eliminating tumor cells.

13. Conclusion

In this dissertation, we introduced a truly unclonable cryptographic key by exploiting resonance energy transfer between networks of fluorophores placed on a nanoscale DNA grid to enable secure, multi-party authentication between legitimate users of the key. We have demonstrated that the RET key has the following important properties:

1. It is infeasible to decipher the underlying physical structure of the key or fabricate a second key when physical access to the first is obtained.
2. It has a complex challenge-response behavior that makes it infeasible to model the behavior of the RET keys.
3. Measurement of all challenge-response pairs on a given key is not possible due to the time-dependent response of the key and its limited lifetime.
4. Determination of all challenge-response pairs on all keys is infeasible due to the large number of CRP's and unique keys.

The RET-key is the first instance where two identical yet unclonable keys can be produced making the two-factor authentication protocol simpler and more secure. Continuous time Markov model results showed changes in the output histograms when minor variations are made to the input or key space. Fluorescence lifetime

measurements were used to confirm the high repeatability of the key signatures and the low levels of noise between identical keys under similar excitation conditions. Significant variations were observed in the key signatures on changing the position of a fluorophore by 2 nm, by modulating the excitation wavelength by only 10 nm or by switching a single fluorophore. Based on experimental results, we estimate that a legitimate user would have an advantage of 10^{340} years over an adversary in arriving at the correct key. This is higher than any of the key's that exist today. In future, an integrated, compact and automatic detection system needs to be developed and techniques to improve the thermal stability of the DNA grids will need to be explored in order to increase the usability of the security system.

This dissertation also introduced DNA self-assembled RET networks as a means to detect secondary and tertiary structure differences between DNA and RNA structures. DNA strands conjugated with fluorophores self-assemble to aberrant structures in the regulatory DNA/RNA regions of genome and form a nanoscale optical network. We found this nanoscale network to be extremely sensitive to molecular variations in the structure. This property enabled us to differentiate between the wild-type structure in normal cells and the aberrant structure in lung and breast cancer cells. Furthermore, the RET network is designed such that an optical signal of a specific wavelength and high intensity is released by the cancerous cell but not by the normal cell. This optical output

can then be used to drive therapeutic agents, such as photosensitizers, which can kill the cancerous cell with high specificity while the normal cell is unharmed. The use of RET networks as a biological sensor several advantages over existing in situ sensors:

1. The nanoscale size of the sensor and the extremely high spatial information density permits its use as a subcellular probe.
2. The use of time-resolved fluorescence detection enables the detection of small variations in the position of fluorophores. This property results in molecular scale spatial resolution and a large number of uniquely identifiable sensors.
3. The sensor makes use of DNA/RNA probe strands, which makes it very specific to the target secondary/tertiary structure (Lin, Jungmann et al. 2012).
4. The optical nature of detection enables rapid *in vitro* and *in vivo* characterization of live cells at picosecond time resolution.
5. The use of oligonucleotides allows access to highly condensed regions in the nucleosome.
6. The characterization technique we use is TCSPC, which is significantly faster and substantially lower in cost compared to imaging. Additionally, since we don't rely on imaging, the sensors we use could potentially be used to study live cells.
7. The RET network is essentially a nanoscale optical computing system. Therefore, if the fluorescence output indicates that a cell is malignant, the output of the RET

network can be used to drive therapeutic agents such as photosensitizers to eliminate the cell.

8. Custom oligonucleotides conjugated with fluorophores can be easily purchased at low cost making the sensor inexpensive. Furthermore, the DNA/RNA self-assembly protocols are not labor intensive or hazardous.

While the RET network shows promise at identifying and treating lung and breast cancer cells with high specificity, extensive *in vitro* and *in vivo* experimental studies are needed to test the specificity of the probe DNA strands to the target structure, improve the stability of the assembled structure in a live cell, control the activation of therapeutic agents, increase the efficacy of the sensor at selectively eliminating cancer cells and reduce any toxic side effects.

Appendix A: Implementation of the Continuous Time Markov Model

```
function main

global Q n

clc
clear A x t

%initial conditions

% Inputs to the script

n = 8; % total number of fluorophores
v = [1 3 4 5 6 8 10 11]'; %list of fluorophores
c = 1; % fluorophores to be excited

ic = zeros(n+n+1,1);
ic(c,1) = 1;
%disp(ic)


r = zeros(n,1);


r =   [0.0881 1.724 -0.544
2.785 -12.586 25.824
-19.58 -15.62 26.376
-12.551 -31.382 28.683
-6.698 -21.328 44.063
-1.059 -22.868 26.994
-10.243 -2.339 16.930
15.386 0.675 9.801
23.344 7.710 5.183
];

  %disp(r)
```

```
%calculation of constants

r0_matrix = importdata('C:\Users\vishwa\Desktop\R0_Matrix.xlsx');
q0_matrix = importdata('C:\Users\vishwa\Desktop\q0_matrix.xlsx');
tau_matrix = importdata('C:\Users\vishwa\Desktop\tau_matrix.xlsx');


d = zeros(n,n);

for aa = 1: n
   for bb = 1:n

      d1 = (r(bb,1)-r(aa,1))*(r(bb,1)-r(aa,1));
      d2 = (r(bb,2)-r(aa,2))*(r(bb,2)-r(aa,2));
      d3 = (r(bb,3)-r(aa,3))*(r(bb,3)-r(aa,3));
      d(aa, bb) = sqrt(d1 + d2 + d3);
      d(aa,bb) = d(aa,bb) *(10^-9);

   end
end

%disp(d)

R = zeros(n,n);

for aa = 1: n
   for bb = 1:n
   R(aa,bb) = r0_matrix.Sheet1(v(aa,1),v(bb,1));
   end
end
%disp(R)

T = zeros(n,1);
k = zeros(n,1);

for aa = 1: n
   T(aa) = tau_matrix.Sheet1(v(aa,1));
   k(aa) = 1/T(aa);
end
```

```
%disp(k)

QY = zeros(n,1);
kqy = zeros(n,1);

for aa = 1: n
   QY(aa) = q0_matrix.Sheet1(v(aa,1));
   kqy(aa) = QY(aa)/T(aa);
end


kret = zeros(n,1);

for aa = 1: n
   for bb = 1:n


    kret(aa,bb) = (1/T(aa))*(R(aa,bb)/d(aa,bb))^6;

   end
end

%disp(kret)

for aa = 1: n
   for bb = 1:n

      if kret(aa,bb) == Inf
         kret(aa,bb) =0;
      end
   end
end


knr = zeros(n,1);

for aa = 1:n
   %knr(aa) = 1/((1/k(aa,1))-((1/kret_total(1,aa) + 1/kqy(aa,1))));
```

```
   knr(aa) = (1/T(aa))-kqy(aa);
end

%disp(knr)


Q = zeros(n+n+1,n+n+1);

for aa = 1:n
   for bb = 1:n

      Q(aa,bb) = kret(aa,bb);

   end
end

c=n+1;

for aa = 1:n
   bb = c;

   Q(aa,bb) = kqy(aa,1);

   c=c+1;
end



for aa = 1:n
   bb = n+n+1;

      Q(aa,bb) = knr(aa,1);

end

for aa = n+1:n+n+1
   for bb = 1:n

      Q(aa,bb) = 0;
   end
```

```
end

for aa = 1:n
   bb =aa;

      Q(aa,bb) = -sum(Q(aa,:));

end



%disp(Q)

%call the solver

trange = [0, 20e-9];

[t,x] = ode45(@TestFunction,trange,ic,[]);



%A = [t,x];
B = x(:,16);
C = t(:,1);



deriv_B =(diff(B))./diff(C);

% final(1,1) =0;
%
% for u = 2: size(deriv_B,1)+1
%
%     final(u,1) = deriv_B(u-1,1);
%
% end


deriv_B(size(t,1), 1) = deriv_B(size(t,1)-1, 1);
```

```
plot(t, deriv_B,'.r');

save('outputs.mat', 'deriv_B');
save('time.mat', 't');


return
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function dxdt = TestFunction(t,x)

global Q n


dxdt = zeros(n+n+1,1);
x = [x(1) ; x(2); x(3); x(4); x(5); x(6); x(7); x(8) ; x(9); x(10); x(11); x(12); x(13); x(14); x(15) ;
x(16); x(17)];
dxdt = Q'*x;


return
```

Appendix B: Calculating the intra-key hamming distance

```
clc
clear global
global n z m  x y s k puf_1 Y X


%File read
%n = input( 'Enter number of files' );
myFolder = 'C:\Users\vishwa\Desktop\survey batch 1 and 2 updated\survey batch 1
488';
%myFolder = 'C:\Users\vishwa\Desktop\survey batch 1 and 2\survey batch 1 and 2
647';
% if ~isdir(myFolder)
%   Message = sprintf('Error: The following folder does not exist:\n%s', myFolder);
%   uiwait(warndlg(Message));
%   return;
% end
filePattern = fullfile(myFolder, '*.asc');
asciiFiles = dir(filePattern);
for k = 1:length(asciiFiles)
  baseFileName = asciiFiles(k).name;
  fullFileName = fullfile(myFolder, baseFileName);
  fprintf(1, 'Now reading %s\n', fullFileName);
   fid=fopen(fullFileName, 'rt');
    header = [fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid)
fgetl(fid) fgetl(fid) fgetl(fid)];
    %disp(header(10:20));
     puf_1(:,k) = fscanf(fid, '%i');
    fclose(fid);
end

n=length(asciiFiles);
sf=50;

% for a = 1:n
% puf_1(:,a) = smooth(puf_1(:,a),20);
```

```
% end

%Image generation 3.05e-12s

% tp = 0.00305:0.00305:0.00305*4096;
% t = tp';


tp = 0.012:0.012:0.012*4096;
t = tp';
Y1 = zeros(4096,1);

for a = 1:n

%     ind = find( puf_1(:,a), 1, 'first');
 puf_1(puf_1 == 0) = 0.001;

    Y1(:,a) = round(log(puf_1(:,a))*sf);
    %Y2(:,a) = round(log(puf_1(:,a))*sf)-min(Y1(:,a));
    Y(:,a) = round(log(puf_1(:,a))*sf)-min(Y1(:,a));


Y(Y == 0) = 1;
X = zeros (4096,max(Y(:,a)));


for b = 1:4095
   for c = 2:max(Y(:,a))
          if Y(b,a) == c
          X(b,c) = (Y(b,a));
          end
   end
end


%gg = nnz(X);
%I = mat2gray(X);
%imshow(mat2gray(X))
```

```
%Hough transform
[H, T, R] = hough(X,'RhoResolution',10,'Theta',-90:0.2:89.8);


%imshow(imadjust(mat2gray(H)),'XData',T,'YData',R,'InitialMagnification', 'fit');
peaks  = houghpeaks(H,1);
%lines = houghlines(X,T,R,peaks);
lines = houghlines(X,T,R,peaks, 'Minlength', 200);

try

   hold on
for k = 1:numel(lines)
   x1 = lines(k).point1(1);
   y1 = lines(k).point1(2);
   x2 = lines(k).point2(1);
   y2 = lines(k).point2(2);
  % plot([x1 x2],[y1 y2],'Color','g','LineWidth', 1)

   ff = -((y2-y1)/(x2-x1))*sf*0.012;
   m(a,k) = ff;

end
hold off

%Calculate lifetimes from semilog

%
%  y =( puf_1(1740:3400,a));
%  x = t(1740:3400,1);
%
%  z = log(y);
%  %plot(z, x)
%
% d = polyfit(x, z, 1);
% s = -1/d(1,1);
% disp (

catch exception
  disp(a);
```

```
 disp(asciiFiles(a).name);
  continue

end
end

save('slopes.mat', 'm');

% mm=zeros(2,2);
%
% for a = 1:n
%     for k=1:2
%     m_mean = mean(m);
%
%     mm(a,k) = (m(a,k)-m_mean(k))*100/m(a,k);
%     end
%
% end



clc
clear global
tic
global  n


string=cell(100,1);
puf_1 = zeros(4096,1);
% %File read
% n = input( 'Enter number of files' );
% for a = 1:n
%    filename = sprintf('%d.asc', a);
%    fid=fopen(filename, 'rt');
%     header = [fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid)
fgetl(fid) fgetl(fid) fgetl(fid)];
%     disp(header(10:27));
%     C(:,a) = textscan(fid, '%s');
%     fclose(fid);
% end
```

```
%File read
myFolder = 'C:\Users\vishwa\Desktop\survey batch 1 and 2 updated\survey batch 1
488';

filePattern = fullfile(myFolder, '*.asc');
asciiFiles = dir(filePattern);
for k = 1:length(asciiFiles)
  baseFileName = asciiFiles(k).name;
  fullFileName = fullfile(myFolder, baseFileName);
  fprintf(1, 'Now reading %s\n', fullFileName);
   fid=fopen(fullFileName, 'rt');
    header = [fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid)
fgetl(fid) fgetl(fid)];
    %disp(header(10:20));
     %puf_1(:,k) = fscanf(fid, '%i');
     C(:,k) = textscan(fid, '%s');
    fclose(fid);
end
n=length(asciiFiles);
%n=2;

Adjustment =zeros(n,1);

%puf_1(all(puf_1==0,2),:)=[];
%Normalize
% puf_1 = normc(puf_1);
% puf_1 = puf_1*1000;
% puf_1 = round(puf_1);

for a = 1:2:(n-1)
    b=a+1;

try
   % for ii=3000:3010
D = str2double(C{a});
E = str2double(C{b});
%E = strcat(D(:,:));
%X = str2double(E);
M1 = mean(D(4101:(end-50),1));
```

```
M2 = mean(E(4101:(end-50),1));

Adjustment(a,1) = (M1-M2)/M1;
Adjustment(b,1) = (M1-M2)/M1;

%Adjustment(a,1) = M1/M2;

%     value = C{a}(ii);
%  patterns(ii,a) = value;
%  patternsnum(ii,a)= str2num(patterns);
  catch exception
  disp(a);
 disp(asciiFiles(a).name);
  continue

end
end
save('fixbleach.mat', 'Adjustment');


toc

crosscorr_raw_lifetime_5_reading_all_files_in_folder;
crosscorr_mcs_bleachingcorr_3;
clc
clear count eee ppp false_negatives final_false_negatives
tic

F = load('fixbleach.mat');
M = load('slopes.mat');

puf_1 = zeros(4096,1);
%File read
%n = input( 'Enter number of files' );
myFolder = 'C:\Users\vishwa\Desktop\survey batch 1 and 2 updated\survey batch 1
488';
% if ~isdir(myFolder)
%   Message = sprintf('Error: The following folder does not exist:\n%s', myFolder);
%   uiwait(warndlg(Message));
%   return;
```

```
% end
filePattern = fullfile(myFolder, '*.asc');
asciiFiles = dir(filePattern);
for k = 1:length(asciiFiles)
  baseFileName = asciiFiles(k).name;
  fullFileName = fullfile(myFolder, baseFileName);
  fprintf(1, 'Now reading %s\n', fullFileName);
   fid=fopen(fullFileName, 'rt');
    header = [fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid)
fgetl(fid) fgetl(fid)];
    %disp(header(10:20));
     puf_1(:,k) = fscanf(fid, '%i');
    fclose(fid);
end

for a = 1:n
puf_1(:,a) = smooth(puf_1(:,a),20);
end


for h = 1:n
  if ((M.m(h,1)) == Inf || (M.m(h,1)) == -Inf || (M.m(h,1)) <0)
     M.m(h,1) =0;
  end

end

n=length(asciiFiles);

n1=2700;
n2=3850;
n3=n2-n1+1;
A= zeros(1800,900,n);
sf = 50;
diff = zeros(n,4096);
count = zeros(n,1);

Y1 = zeros(4096,1);
Y = zeros(4096,1);
```

```
for a=1:n

   puf_1(puf_1 == 0) = 0.1;

%    Z1(1:n3,a) = round(log(puf_1(n1:n2,a))*sf);
%    Z(1:n3,a) = round(log(puf_1(n1:n2,a))*sf)-min(Z1(1:n3,a));

if ((mod(a,2) ~= 0) && (F.Adjustment(a,1) <0))
%number is odd

   Y1(1:n3,a) = round(log((puf_1(n1:n2,a))+ (puf_1(n1:n2,a)*-F.Adjustment(a,1)))*sf);
   Y(1:n3,a) = round(log((puf_1(n1:n2,a))+ (puf_1(n1:n2,a)*-F.Adjustment(a,1)))*sf)-
min(Y1(1:n3,a));

elseif ((mod(a,2) == 0) && (F.Adjustment(a,1) >= 0))
%number is even

   Y1(1:n3,a) = round(log((puf_1(n1:n2,a))+ (puf_1(n1:n2,a)*F.Adjustment(a,1)))*sf);
   Y(1:n3,a) = round(log((puf_1(n1:n2,a))+ (puf_1(n1:n2,a)*F.Adjustment(a,1)))*sf)-
min(Y1(1:n3,a));

else

   Y1(1:n3,a) = round(log(puf_1(n1:n2,a))*sf);
   Y(1:n3,a) = round(log(puf_1(n1:n2,a))*sf)-min(Y1(1:n3,a));

%    Y1(1:n3,a) = round(log(puf_1(n1:n2,a))*sf*Adjustment(a,1));
%    Y(1:n3,a) = round(log(puf_1(n1:n2,a))*sf*Adjustment(a,1))-min(Y1(1:n3,a));

%    Y1(1:n3,a) = round(log(puf_1(n1:n2,a))*sf);
%    Y(1:n3,a) = round(log(puf_1(n1:n2,a))*sf)-min(Y1(1:n3,a));

end

Y1(Y1 == 0) = 0.001;
X = zeros (4096,400);
%BW = zeros (4096,400);

for b = 1:4094
```

```
   for c = 2:max(Y1(:,a))
         if Y1(b,a) == c
         X(b,c) = Y1(b,a);
         end
   end
end

Y1(Y1 == 0) = 0.001;
X = zeros (4096,400);


for b = 1:4094
   for c = 2:max(Y1(:,a))
         if Y1(b,a) == c
         X(b,c) = Y1(b,a);
         end
   end
end


[H, T, R] = hough(X,'RhoResolution',5,'Theta',-90:0.2:89.8);
% B(:,:) = A(:,:,1);
% C(:,:) = A(:,:,2);
%disp(nnz(A(:,:,a)))
H(1800,900) =0;

A(:,:,a)=  H;
% B(:,:) = A(:,:,1);
% C(:,:) = A(:,:,2);
% D(:,:) = A(:,:,3);
end

 hamming_dist = zeros(n,1);
 hamming_dist_perc = zeros(n,1);

for a=1:1:n-25
 xxx=a+24;

if ((M.m(a,1))>2)
```

```
mmm1 = (round(90-atand (M.m(a,1))))*5;
% disp(M.m(a,1))
% disp(mmm1)
iii1=mmm1-11;
iii2=mmm1+11;


eee = zeros (1800,(iii2-iii1+1)*2);

for iii = iii1:iii2
   hhh =1;
for ggg = 1:1600

   if A(ggg, iii,a) || A(ggg, iii,xxx) ~= 0
   eee(hhh,iii) = A(ggg,iii,a);
   eee(hhh,iii+(iii2-iii1+1)) = A(ggg,iii,xxx);
   hhh = hhh+1;
   end

end
end

%disp(nnz(eee))
eee(all(eee==0,2),:)=[];
eee = eee(:,any(eee));

eee(eee <= 9) = 0;
eee(eee > 9) = 1;

hamming_dist_1 = 0;
for j = 1:(iii2-iii1+1)
   hd1 = sum(eee(:,j)~=eee(:,(j+(iii2-iii1+1))));
   hamming_dist_1 = hamming_dist_1 + hd1;
end

hamming_dist_1_perc = (hamming_dist_1 *100)/(size(eee,1)*23);

% for qq=1:size(eee,1)
%
%
```

```
%    if eee(qq,1)>=9
%     diff = ((eee(qq,1)- eee(qq,2))*100)/(eee(qq,1));
%     if (diff >= 10 || diff <= -10)
%         count(a) = count(a) +1;
%     end
%    end
% end



mmm2 = (round(90-atand (M.m(xxx,1))))*5;
% disp(M.m(a,1))
% disp(mmm1)
iii1=mmm1-11;
iii2=mmm1+11;

ppp = zeros (1800,(iii2-iii1+1)*2);

for iii = iii1:iii2
   hhh =1;
for ggg = 1:1600

   if A(ggg, iii,a) || A(ggg, iii,xxx) ~= 0
   ppp(hhh,iii) = A(ggg,iii,a);
   ppp(hhh,iii+(iii2-iii1+1)) = A(ggg,iii,xxx);
   hhh = hhh+1;
   end

end
end


ppp(all(ppp==0,2),:)=[];
ppp = ppp(:,any(ppp));

ppp(ppp <= 9) = 0;
ppp(ppp > 9) = 1;

hamming_dist_2 = 0;
for j = 1:(iii2-iii1+1)
```

```
   hd2 = sum(ppp(:,j)~=ppp(:,(j+(iii2-iii1+1))));
   hamming_dist_2 = hamming_dist_2 + hd2;
end

hamming_dist_2_perc = (hamming_dist_2 *100)/(size(ppp,1)*23);

% for qq=1:size(ppp,1)
%    if ppp(qq,1)>=9
%     diff = ((ppp(qq,1)- ppp(qq,2))*100)/(ppp(qq,1));
%     if (diff >= 10 || diff <= -10)
%         count(a) = count(a) +1;
%     end
%    end
% end

hamming_dist(a) = hamming_dist_1 + hamming_dist_2;
hamming_dist_perc(a) = hamming_dist_1_perc + hamming_dist_2_perc;

end
end


false_negatives_100 =0;
false_negatives_200 =0;
false_negatives_300 =0;
false_negatives_400 =0;
false_negatives_500 =0;

colliding_samples = zeros(5,1);
v=1;

for a =1:n
   if hamming_dist(a) >=100

       colliding_samples(v,5) = a;
       colliding_samples(v,6) = xxx;
       colliding_samples(v,7) = xxx-a;


       v=v+1;
```

```
        false_negatives_100 = false_negatives_100 + 1;
    end
end

final_false_negatives_100 = (false_negatives_100*100)/(n/2);

v=1;
for a =1:n
    if hamming_dist(a) >=200
        colliding_samples(v,9) = a;
        colliding_samples(v,10) = xxx;
        colliding_samples(v,11) = xxx-a;

        v=v+1;

        false_negatives_200 = false_negatives_200 + 1;
    end
end

final_false_negatives_200 = (false_negatives_200*100)/(n/2);

v=1;
for a =1:n
    if hamming_dist(a) >=300
        colliding_samples(v,12) = a;
        colliding_samples(v,13) = xxx;
        colliding_samples(v,14) = xxx-a;

        v=v+1;

        false_negatives_300 = false_negatives_300 + 1;
    end
end

final_false_negatives_300 = (false_negatives_300*100)/(n/2);

v=1;
for a =1:n
    if hamming_dist(a) >=400
        colliding_samples(v,1) = a;
```

```
        colliding_samples(v,2) = xxx;
        colliding_samples(v,3) = xxx-a;

        v=v+1;

        false_negatives_400 = false_negatives_400 + 1;
    end
end

final_false_negatives_400 = (false_negatives_400*100)/(n/2);

v=1;

for a =1:n
    if hamming_dist(a) >=500

        colliding_samples(v,15) = a;
        colliding_samples(v,16) = xxx;
        colliding_samples(v,17) = xxx-a;

        v=v+1;

        false_negatives_500 = false_negatives_500 + 1;
    end
end

final_false_negatives_500 = (false_negatives_500*100)/(n/2);

toc
```

Appendix C: Calculating the inter-key hamming distance

```
clc
clear global
global n z m  x y s k puf_1 Y X


%File read
%n = input( 'Enter number of files' );
myFolder = 'C:\Users\vishwa\Desktop\survey batch 1 and 2 updated\survey batch 1
488';
%myFolder = 'C:\Users\vishwa\Desktop\survey batch 1 and 2\survey batch 1 and 2
647';
% if ~isdir(myFolder)
%   Message = sprintf('Error: The following folder does not exist:\n%s', myFolder);
%   uiwait(warndlg(Message));
%   return;
% end
filePattern = fullfile(myFolder, '*.asc');
asciiFiles = dir(filePattern);
for k = 1:length(asciiFiles)
  baseFileName = asciiFiles(k).name;
  fullFileName = fullfile(myFolder, baseFileName);
  fprintf(1, 'Now reading %s\n', fullFileName);
   fid=fopen(fullFileName, 'rt');
    header = [fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid)
fgetl(fid) fgetl(fid) fgetl(fid)];
    %disp(header(10:20));
     puf_1(:,k) = fscanf(fid, '%i');
    fclose(fid);
end

n=length(asciiFiles);
sf=50;

% for a = 1:n
% puf_1(:,a) = smooth(puf_1(:,a),20);
% end
```

```
%Image generation 3.05e-12s

% tp = 0.00305:0.00305:0.00305*4096;
% t = tp';


tp = 0.012:0.012:0.012*4096;
t = tp';
Y1 = zeros(4096,1);

for a = 1:n

%     ind = find( puf_1(:,a), 1, 'first');
 puf_1(puf_1 == 0) = 0.001;

    Y1(:,a) = round(log(puf_1(:,a))*sf);
    %Y2(:,a) = round(log(puf_1(:,a))*sf)-min(Y1(:,a));
    Y(:,a) = round(log(puf_1(:,a))*sf)-min(Y1(:,a));


Y(Y == 0) = 1;
X = zeros (4096,max(Y(:,a)));


for b = 1:4095
   for c = 2:max(Y(:,a))
         if Y(b,a) == c
         X(b,c) = (Y(b,a));
         end
   end
end


%gg = nnz(X);
%I = mat2gray(X);
%imshow(mat2gray(X))


%Hough transform
```

```
[H, T, R] = hough(X,'RhoResolution',10,'Theta',-90:0.2:89.8);


%imshow(imadjust(mat2gray(H)),'XData',T,'YData',R,'InitialMagnification', 'fit');
peaks  = houghpeaks(H,1);
%lines = houghlines(X,T,R,peaks);
lines = houghlines(X,T,R,peaks, 'Minlength', 200);

try

   hold on
for k = 1:numel(lines)
   x1 = lines(k).point1(1);
   y1 = lines(k).point1(2);
   x2 = lines(k).point2(1);
   y2 = lines(k).point2(2);
  % plot([x1 x2],[y1 y2],'Color','g','LineWidth', 1)

   ff = -((y2-y1)/(x2-x1))*sf*0.012;
   m(a,k) = ff;

end
hold off

%Calculate lifetimes from semilog

%
%  y =( puf_1(1740:3400,a));
%  x = t(1740:3400,1);
%
%  z = log(y);
%  %plot(z, x)
%
% d = polyfit(x, z, 1);
% s = -1/d(1,1);
% disp (

catch exception
  disp(a);
 disp(asciiFiles(a).name);
```

```
 continue

end
end

save('slopes.mat', 'm');

% mm=zeros(2,2);
%
% for a = 1:n
%     for k=1:2
%     m_mean = mean(m);
%
%     mm(a,k) = (m(a,k)-m_mean(k))*100/m(a,k);
%     end
%
% end



clc
clear global
tic
global  n


string=cell(100,1);
puf_1 = zeros(4096,1);
% %File read
% n = input( 'Enter number of files' );
% for a = 1:n
%    filename = sprintf('%d.asc', a);
%    fid=fopen(filename, 'rt');
%     header = [fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid)
fgetl(fid) fgetl(fid) fgetl(fid)];
%     disp(header(10:27));
%     C(:,a) = textscan(fid, '%s');
%     fclose(fid);
% end
```

```
%File read
myFolder = 'C:\Users\vishwa\Desktop\survey batch 1 and 2 updated\survey batch 1
488';

filePattern = fullfile(myFolder, '*.asc');
asciiFiles = dir(filePattern);
for k = 1:length(asciiFiles)
  baseFileName = asciiFiles(k).name;
  fullFileName = fullfile(myFolder, baseFileName);
  fprintf(1, 'Now reading %s\n', fullFileName);
   fid=fopen(fullFileName, 'rt');
    header = [fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid)
fgetl(fid) fgetl(fid)];
    %disp(header(10:20));
     %puf_1(:,k) = fscanf(fid, '%i');
     C(:,k) = textscan(fid, '%s');
    fclose(fid);
end
n=length(asciiFiles);
%n=2;

Adjustment =zeros(n,1);
T =zeros(n,1);

%puf_1(all(puf_1==0,2),:)=[];
%Normalize
% puf_1 = normc(puf_1);
% puf_1 = puf_1*1000;
% puf_1 = round(puf_1);

for a = 1:n
    for xxx =a+1:n

try
   % for ii=3000:3010


D = str2double(C{a});
E = str2double(C{xxx});
```

```
T(a) = mean(D(4101:(end-50),1));
%E = strcat(D(:,:));
%X = str2double(E);
M1 = mean(D(4101:(end-50),1));
M2 = mean(E(4101:(end-50),1));

Adjustment(a,xxx) = (M1-M2)/M1;
%Adjustment(b,1) = (M1-M2)/M1;

%Adjustment(a,1) = M1/M2;



%     value = C{a}(ii);
%  patterns(ii,a) = value;
%  patternsnum(ii,a)= str2num(patterns);
  catch exception
  disp(a);
 disp(asciiFiles(a).name);
  continue

end
    end
end
save('fixbleach.mat', 'Adjustment');


toc

crosscorr_raw_lifetime_5_reading_all_files_in_folder;
crosscorr_mcs_bleachingcorr_3_interpuf;


clc
clear count eee ppp false_positives final_false_positives colliding_samples
tic

F = load('fixbleach.mat');
M = load('slopes.mat');
```

```
puf_1 = zeros(4096,1);
%File read
%n = input( 'Enter number of files' );
myFolder = 'C:\Users\vishwa\Desktop\survey batch 1 and 2 updated\survey batch 1
647';
% if ~isdir(myFolder)
%   Message = sprintf('Error: The following folder does not exist:\n%s', myFolder);
%   uiwait(warndlg(Message));
%   return;
% end
filePattern = fullfile(myFolder, '*.asc');
asciiFiles = dir(filePattern);
for k = 1:length(asciiFiles)
  baseFileName = asciiFiles(k).name;
  fullFileName = fullfile(myFolder, baseFileName);
  fprintf(1, 'Now reading %s\n', fullFileName);
   fid=fopen(fullFileName, 'rt');
    header = [fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid) fgetl(fid)
fgetl(fid) fgetl(fid) fgetl(fid)];
    %disp(header(10:20));
     puf_1(:,k) = fscanf(fid, '%i');
    fclose(fid);
end

n=length(asciiFiles);

for h = 1:n
   if ((M.m(h,1)) == Inf || (M.m(h,1)) == -Inf || (M.m(h,1)) <0)
      M.m(h,1) =0;
   end

end

n1=2700;
n2=3850;
n3=n2-n1+1;
A= zeros(1800,900,n);
sf = 50;
diff = zeros(n,4096);
count = zeros(n,1);
```

```
Y1 = zeros(4096,1);
Y = zeros(4096,1);

W1 = zeros(4096,1);
W = zeros(4096,1);

 hamming_dist = zeros(n,n);
 hamming_dist_perc = zeros(n,n);

% for a = 1:n
% puf_1(:,a) = smooth(puf_1(:,a),20);
% end

 puf_1(puf_1 == 0) = 0.1;


for a=1:n
 for xxx = (a+1):n



Y1(1:n3,a) = round(log((puf_1(n1:n2,a))+ (puf_1(n1:n2,a)*-Adjustment(a,xxx)))*sf);
Y(1:n3,a) = round(log((puf_1(n1:n2,a))+ (puf_1(n1:n2,a)*-Adjustment(a,xxx)))*sf)-
min(Y1(1:n3,a));


Y1(Y1 == 0) = 0.001;
X = zeros (4096,400);


for b = 1:4094
   for c = 2:max(Y1(:,a))
         if Y1(b,a) == c
         X(b,c) = Y1(b,a);
         end
   end
end
```

```
[H, T, R] = hough(X,'RhoResolution',5,'Theta',-90:0.2:89.8);



W1(1:n3,a) = round(log(puf_1(n1:n2,a))*sf);
W(1:n3,a) = round(log(puf_1(n1:n2,a))*sf)-min(Y1(1:n3,a));



W1(W1 == 0) = 0.001;
Z = zeros (4096,400);


for b = 1:4094
   for c = 2:max(W1(:,a))
         if W1(b,a) == c
         Z(b,c) = W1(b,a);
         end
   end
end


[H1, T, R] = hough(Z,'RhoResolution',5,'Theta',-90:0.2:89.8);




if ((M.m(a,1))>2)

mmm1 = (round(90-atand (M.m(a,1))))*5;
% disp(M.m(a,1))
% disp(mmm1)
iii1=mmm1-11;
iii2=mmm1+11;

eee = zeros (1800,(iii2-iii1+1)*2);

for iii = iii1:iii2
   hhh =1;
for ggg = 1:1600
```

```
   if H(ggg, iii) || H1(ggg, iii) ~= 0
   eee(hhh,iii) = H(ggg,iii);
   eee(hhh,iii+(iii2-iii1+1)) = H1(ggg,iii);
   hhh = hhh+1;
   end

end
end

%disp(nnz(eee))
eee(all(eee==0,2),:)=[];
eee = eee(:,any(eee));

eee(all(eee==0,2),:)=[];
eee = eee(:,any(eee));

eee(eee <= 9) = 0;
eee(eee > 9) = 1;

hamming_dist_1 = 0;
for j = 1:(iii2-iii1+1)
   hd1 = sum(eee(:,j)~=eee(:,(j+(iii2-iii1+1))));
   hamming_dist_1 = hamming_dist_1 + hd1;
end

hamming_dist_1_perc = (hamming_dist_1 *100)/(size(eee,1)*23);


mmm2 = (round(90-atand (M.m(xxx,1))))*5;
% disp(M.m(a,1))
% disp(mmm1)
iii1=mmm2-11;
iii2=mmm2+11;

ppp = zeros (1800,(iii2-iii1+1)*2);

for iii = iii1:iii2
   hhh =1;
for ggg = 1:1600
```

```
    if H(ggg, iii) || H1(ggg, iii) ~= 0
    ppp(hhh,iii) = H(ggg,iii);
    ppp(hhh,iii+(iii2-iii1+1)) = H1(ggg,iii);
    hhh = hhh+1;
    end

end
end


ppp(all(ppp==0,2),:)=[];
ppp = ppp(:,any(ppp));

ppp(ppp <= 9) = 0;
ppp(ppp > 9) = 1;

hamming_dist_2 = 0;
for j = 1:(iii2-iii1+1)
    hd2 = sum(ppp(:,j)~=ppp(:,(j+(iii2-iii1+1))));
    hamming_dist_2 = hamming_dist_2 + hd2;
end

hamming_dist_2_perc = (hamming_dist_2 *100)/(size(ppp,1)*23);

hamming_dist(a,xxx) = hamming_dist_1 + hamming_dist_2;
hamming_dist_perc(a,xxx) = hamming_dist_1_perc + hamming_dist_2_perc;

end

 end
end

false_positives_400 =0;
false_positives_100 =0;
false_positives_200 =0;
false_positives_300 =0;
false_positives_500 =0;

colliding_samples = zeros(5,1);
```

```
v=1;
hamming_dist(hamming_dist == 0) = NaN;

for a =1:n
   for xxx = 1:n
   if hamming_dist(a,xxx) <=400
      colliding_samples(v,1) = a;
      colliding_samples(v,2) = xxx;
      colliding_samples(v,3) = xxx-a;
      false_positives_400 = false_positives_400 + 1;
      v=v+1;
   end
   end
end

final_collisions_400 = ((false_positives_400*100)/nnz(hamming_dist));
%Interpuf_bit_flips_400 = sum(hamming_dist_perc(:))/nnz(hamming_dist_perc);

v=1;
for a =1:n
   for xxx = 1:n
   if hamming_dist(a,xxx) <=100
      colliding_samples(v,5) = a;
      colliding_samples(v,6) = xxx;
      colliding_samples(v,7) = xxx-a;
      false_positives_100 = false_positives_100 + 1;
      v=v+1;
   end
   end
end

final_collisions_100 = ((false_positives_100*100)/nnz(hamming_dist));
%Interpuf_bit_flips = sum(hamming_dist_perc(:))/nnz(hamming_dist_perc);

v=1;
for a =1:n
   for xxx = 1:n
   if hamming_dist(a,xxx) <=200
      colliding_samples(v,9) = a;
      colliding_samples(v,10) = xxx;
```

```
        colliding_samples(v,11) = xxx-a;
        false_positives_200 = false_positives_200 + 1;
        v=v+1;
    end
    end
end

final_collisions_200 = ((false_positives_200*100)/nnz(hamming_dist));
%Interpuf_bit_flips = sum(hamming_dist_perc(:))/nnz(hamming_dist_perc);

v=1;
for a =1:n
    for xxx = 1:n
    if hamming_dist(a,xxx) <=300
        colliding_samples(v,12) = a;
        colliding_samples(v,13) = xxx;
        colliding_samples(v,14) = xxx-a;
        false_positives_300 = false_positives_300 + 1;
        v=v+1;
    end
    end
end

final_collisions_300 = ((false_positives_300*100)/nnz(hamming_dist));
%Interpuf_bit_flips = sum(hamming_dist_perc(:))/nnz(hamming_dist_perc);

v=1;
for a =1:n
    for xxx = 1:n
    if hamming_dist(a,xxx) <=500
        colliding_samples(v,15) = a;
        colliding_samples(v,16) = xxx;
        colliding_samples(v,17) = xxx-a;
        false_positives_500 = false_positives_500 + 1;
        v=v+1;
    end
    end
end

final_collisions_500 = ((false_positives_500*100)/nnz(hamming_dist));
```

```
%Interpuf_bit_flips = sum(hamming_dist_perc(:))/nnz(hamming_dist_perc);



binsize = 1:1:n;
bincolumn1 = hist(colliding_samples(:,1),binsize);
bincolumn1 = bincolumn1';
bincolumn2 = hist(colliding_samples(:,2),binsize);
bincolumn2 = bincolumn2';
bincolumn = (bincolumn1+bincolumn2)/2;

toc
```

Appendix D: Implementation of the maximum entropy algorithm

```
clc;
clear Q prob e I nn final h xx yy


a=1;

for n=a:500;

Q = zeros(n,n);


for iii=1:n
   for jjj = iii: (iii+13-1)
   if jjj<=n
      Q(iii,jjj) = 1/15;
   end
   end
end




for k = 1:n

   if k ==1
      e(k,1)=1;
   else
   e(k,1)= 0;
   end

end
```

```
I = eye(n);
xx= (I-Q');
yy= inv(I-Q');
nn = (inv(I - Q'))*e;

entropy =0;
h=zeros(n,1);

for iii = a:n
   for jjj = a:n

      if ((Q(iii,jjj)~=0 ))
      h(iii) = h(iii) + (Q(iii,jjj)*log2(Q(iii,jjj)));
      end

   end
end

% if n>22
% for iii = (n-22):n
%     h(iii) = -4.054;
%
% end
% end


for iii = a:n

       entropy = entropy + (nn(iii,1)* -h(iii));

end

%entropy = -entropy;
final(n) = entropy;
%final_corrected(n) = entropy - ((n-1)/2);
%final_corrected(n) = entropy - (((10^4)*n)/(2*0.63*(10^7)));
ME= final';
%ME_corrected = final_corrected';
```

```
end
```

Appendix E: Implementation of the average entropy algorithm

```
clc;
clear Q prob e I nn final h xx yy bincolumn_ae

totalnodes = 500;
totaltrials = 250;
interationsize =20;
binsize_me = 0.01:0.01:1;
final = zeros(1,100);
finalTE = zeros(100,1);
All_entropy = zeros(totaltrials, totalnodes);

for a=3:totalnodes

%     bincolumn_me(a,1:100) = zeros(1,100);
%     bincolumn_ae(1:100,a) = zeros(100,1);
    %allTE = zeros(1000,1);
    %u=1;
    sumentropy = 0;

    for trials =1:totaltrials

Q = zeros(a,a);

TEscale = 0:0.01:1;
TEscale2 = 0.81:0.01:1;

for iii=1:a

TE = zeros(20,1);
sum = 0;

     for jjj = iii: (iii+interationsize)
```

```
    if jjj<=a

       random = randi(101, 1);
       TE(jjj) = TEscale(random);
       sum = sum + TE(jjj);


    end
     end


     for jjj = iii: (iii+interationsize)

    if jjj<=a

       random = randi(20, 1);
       r = TEscale2(random);
       TE(jjj) = (TE(jjj)/sum)*r;
       %allTE(u) =(TE(jjj));
       %u =u+1;
    end
     end


        for jjj = iii+1: (iii+interationsize)

    if jjj<=a

       Q(iii,jjj) = TE(jjj);


    end
        end

end



for k = 1:a
```

```
   if k ==1
      e(k,1)=1;
   else
   e(k,1)= 0;
   end
end

I = eye(a);
ww = Q';
xx= (I-Q');

yy= inv(I-Q');
nn = (inv(I - Q'))*e;

entropy =0;
h=zeros(a,1);

for iii = 1:a
   for jjj = 1:a

      if ((Q(iii,jjj)~=0 ))
      h(iii) = h(iii) + (Q(iii,jjj)*log2(Q(iii,jjj)));
      end

   end
end


for iii = 1:(a)

       entropy = entropy + (nn(iii,1)* -h(iii));
       %entropy = entropy - h(iii);

end
All_entropy(trials,a) = entropy;

sumentropy = sumentropy + entropy;
%disp(sumentropy)
   end
```

```
final(a) = sumentropy/totaltrials;


% bincolumn_me = hist(allTE,binsize_me);
% bincolumn_ae(1:100,a) = bincolumn_me';

end


ME= final';
% All_entropy = round(All_entropy*100000);
% t=0;
%
% for vvv = 1:(totaltrials-1)
%     for www = vvv+1:totaltrials
%
%         if All_entropy(vvv,totalnodes) == All_entropy(www,totalnodes)
%             disp(vvv)
%             disp(www)
%             t=t+1;
%         end
%     end
% end

% for l = 1:100
%     sum =0;
%     for m = 20:totalnodes
%
% sum = sum + (bincolumn_ae(l,m));
%
%     end
%
%     finalTE(l) = sum;
%
% end
```

References

B. Skoric, S. M. T. K. and P. Tuyls (2006). Information-theoretic analysis of capacitive physical unclonable functions, J. Appl. Phys.

Barhoumi, A. and N. J. Halas (2010). "Surface-Enhanced Raman Spectroscopy of DNA." JACS **132**.

Barrett, C. W., et al. (2013). "Tumor suppressor function of the plasma glutathione peroxidase Gpx3 in colitis-associated carcinoma." Cancer Research **73**(3).

Becker, W. (2008). The Bh TCSPC handbook, Becker & Hickl GmbH.

Bhatia, D., et. al. (2011). "A synthetic icosahedral DNA-based host–cargo complex for functional in vivo imaging." Nature Communications **2**.

Bosch, C., et al. (2008). "Efficient Helper Data Key Extractor on FPGAs." Lecture Notes in Computer Science **5154**: 181-197.

Buckhout-White, S., et al. (2013). "TEM imaging of unstained DNA nanostructures using suspended graphene." Soft Matter **9**.

Bushnell, M., & Agrawal, V (2000). Essentials of electronic testing for digital, memory, and mixed-signal VLSI circuits, Springer.

Cantor, C. R. and L. S. Cassandra (2004). Genomics: The Science and Technology Behind the Human Genome Project, John Wiley & Sons.

Caspers, P. J., et al. (2003). "Combined in vivo confocal raman spectroscopy and confocal microscopy of human skin." Biophysical Journal **85**(1).

Chok, N. S. (2010). Pearson's Versus Spearman's and Kendall's Correlation Coefficients for Continuous Data, University of Pittsburgh. **Master's Thesis**.

Clemens, H. and B. Christian (2013). "Cloning Physically Unclonable Functions." 2013 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST) **6**: 1-6.

Coldren, L., et al. (2012). Diode Lasers and Photonic Integrated Circuits, John Wiley and Sons.

Cover, T. T., Joy (1991). Elements of Information Theory, Wiley Series in Telecommunications and Signal Processing.

D.G. Porter, J. A. O. S. R. S. I. and M. W. Muller (1994). Physically based information science of magnetic recording: II. Physical sources of medium noise, IEEE Trans. on Magn. **30**.

Dillon, L., W, et al. (2013). "Role of DNA secondary structures in fragile site breakage along human chromosome 10." Human Molecular Genetics.

Duda, R. O. and P. Hart, E. (1972). "Use of the Hough Transformation to Detect Lines and Curves in Pictures." Communications of the ACM **15**.

Erie, D., et al. (1987). "A dumbbell-shaped double-hairpin structure of DNA a thermodynamic investigation." Biochemistry **26**.

Freiden, M., et al. (2003). "Nuclease stability of LNA oligonucleotides and LNA-DNA chimeras." Nucleosides Nucleotides Nucleic Acids **22**.

Giesen, U., et al. (1998). "A formula for thermal stability (Tm) prediction of PNA/DNA duplexes." Nucl. Acids Res. **26**.

Goldreich, O. (2001). Foundations of Cryptography, Volume 1, Basic Tools, Cambridge University Press.

Gordon, M. P., et al. (2004). "Single-molecule high-resolution imaging with photobleaching." Proceedings of the National Academy of Sciences of the United States of America **101**(17): 6462-6465.

Guajardo, J., et al. (2007). "FPGA Intrinsic PUFs and Their Use for IP Protection." Proceedings of the 9th international workshop on Cryptographic Hardware and Embedded Systems.

Han, D., et al. (2011). "DNA Origami with Complex Curvatures in Three-dimensional Space." Science **332**.

Herder, C., et al. (2014). "Physical Unclonable Functions and Applications: A Tutorial." Proceedings of the IEEE **102**(8).

Hilbert, M. and P. López (2011). "The World's Technological Capacity to Store, Communicate, and Compute Information." Science **332**(6025): 60-65.

Huang, B., et al. (2008). "Three-Dimensional Super-Resolution Imaging by Stochastic Optical Reconstruction Microscopy." Science **319**(5864): 810-813.

Humphrey, W., et al. (1996). "VMD - Visual Molecular Dynamics." J. Molec. Graphics **14**.

Kertesz, M., et al. (2007). "The role of site accessibility in microRNA target recognition." Natue Genet. **39**.

Kevin, A. W., et al. (2006). "Selective 2¢-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution." Nature Protocols **1**(3).

King, M. C., et al. (2003). "Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2." Science **302**.

Lakowicz, J. R. (1999). Principles of Fluorescence Spectroscopy. New York, Kluwer Academic / Plenum Publishers.

Lee, H., et. al. (2012). "Molecularly self-assembled nucleic acid nanoparticles for targeted in vivo siRNA delivery." Nature Nanotechnology **7**.

Lee, J. W., et al. (2004). A technique to build a secret key in integrated circuits for identification and authentication applications, VLSI Circuits, 2004. Digest of Technical Papers**:** 176-179.

Leontis, B. N. (2012). RNA 3D structure analysis and prediction, Springer.

LifeTechnologies (2013). Fluorescence SpectraViewer.

Lin, C., et al. (2012). "Submicrometre geometrically encoded fluorescent barcodes self-assembled from DNA." Nat Chem **4**(10): 832-839.

Lin, C., et al. (2008). "In vivo cloning of artifical DNA nanostructures." Pnas **105**(46).

Lorenz, R., et al. (2011). "ViennaRNA Package 2.0." Algorithms for Molecular Biology **6**(26).

Macke, T. and D. A. Case (1997). Modeling unusual nucleic acid structures. Molecular Modeling of Nucleic Acids, ACS Symposium Series. **682**.

Magnus, L., et al. (2002). "Lead(II) as a probe for investigating RNA structure in vivo." RNA **8**.

MicrosoftCorporation (2013). "Understanding and Evaluating Virtual Smart Cards." (1).

Mortimer, S., A, et al. (2014). "Insights into RNA structure and function from genome-wide studies." Nature Reviews Genetics **15**.

Mousa, A. S. and J. D. Bharali (2011). "Nanotechnology-Based Detection and Targeted Therapy in Cancer: Nano-Bio Paradigms and Applications." Cancers **3**.

Mura, S., et al. (2013). "Stimuli-responsive nanocarriers for drug delivery." Nature Materials **12**.

Pappu, R. (2001). Physical One-Way Functions.

Pappu, R., et al. (2002). "Physical One-Way Functions." Science **297**(5589): 2026-2030.

Pappu, R., et al. (2002). "Physical One-way Functions." Science **297**(5589): 2026-2030.

Piper, S. K., et al. (2013). "Towards whole-body fluorescence imaging in humans." PLoS ONE **8**(12).

Pistol, C., et al. (2010). "Encoded multi-chromophore response for simulataneous label-free detection." Small **6**(7): 843-850.

Popenda, M., et al. (2006). "High-throughput method for the prediction of low-resolution, three-dimensional RNA structures." Nucleic Acids Symp Ser (Oxf) **50**.

Popenda, M., et al. (2011). "RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures." BMC Bioinformatics **11**(231).

Rittweger, E., et al. (2009). "STED microscopy reveals crystal colour centres with nanometric resolution." Nat Photon **3**(3): 144-147.

Rouskin, S., et. al. (2013). "Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo." Nano Lett. **505**.

Ruhrmair, U., et al. (2013). "Power and Timing Side Channels for PUFs and their Efficient Exploitation." Crypology ePrint Archive 2013/851.

Rust, M. J., et al. (2006). "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)." Nat Meth **3**(10): 793-796.

Sabarinathan, R., et al. (2014). "Transcriptome-Wide Analysis of UTRs in Non-Small Cell Lung Cancer Reveals Cancer-Related Genes with SNV Induced Changes on RNA Secondary Structure and miRNA Target Sites." PLoS ONE **9**(1).

Saerens M, A. Y., Fouss F, Yen L. (2009). "Randomized shortest-path problems: two related models." Neural Comput.

Sailesh, G., et al. (2007). "Implication of BRCA2 -26G>A 5' untranslated region polymorphism in susceptibility to sporadic breast cancer and its modulation by p53 codon 72 Arg>Pro polymorphism." Breast Cancer Research **9**(5).

Sandra, E. W., et al. (2000). "Use of dimethylsulfate to probe RNA structure in vivo." Methods in Enzymology **318**.

Shcherbakova, I. and M. Brenowitz (2008). "Monitoring structural changes in nucleic acids with single residue spatial and millisecond time resolution by quantitative hydroxyl radical footprinting." Nature Protocols **3**(2).

Shigetaka, N. and F. Kenzo (2014). "Creation of DNA array structure equipped with heat resistance by ultrafast photocrosslinking." J Chem Technol Biotechnol **89**.

Shih, W. M., et al. (2004). "A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron." Nature **427**.

Shtengel, G., et al. (2009). "Interferometric fluorescent super-resolution microscopy resolves 3D cellular ultrastructure." Proceedings of the National Academy of Sciences.

Snedecor, G. W. and W. G. Cochran (1989). Statistical Methods, Iowa State University Press.

Sorensen, J. J., et al. (2004). "Solution Structure of a Dsdna : Lna Triplex." Nucleic Acids Research **32**(20): 6078-6085.

Stanzione, S. and G. Iannaccone (2009). "Silicon physical unclonable function resistant to a 10 -trial brute force attack in 90 nm CMOS." 116–117.

Stefanie, A., Mortimer. and M. Kevin, Weeks. "Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution." Nature Protocols **4**(10).

Stinson, D. (2006). Cryptography Theory and Practise, Chapman & Hall/CRC.

Stoneley, M. and A. E. Willis (2003). "Aberrant regulation of translation initiation in tumorignesis." Curr Mol Med **3**(7).

Suh, G. E. and S. Devadas (2007). "Physical unclonable functions for device authentication and secret key generation." 9–14.

Trivedi, K. S. (2001). Probability and Statistics with Reliability, Queuing, and Computer Science Applications, John Wiley and Sons, New York.

Tucker, B. J. and R. R. Breaker (2005). "Riboswitches as versatile gene control elements." Curr. Opin. Struct. Biol. **15**.

Usuda, J. M., et al. (2006). "Photodynamic Therapy (PDT) for Lung Cancers." Journal of Thoracic Oncology **1**(5).

Valoczi, A., et al. (2004). "Sensitive and Specific Detection of Micrornas by Northern Blot Analysis Using Lna-modified Oligonucleotide Probes." Nucleic Acids Research **32**(22 e175): -.

Vandersypen, L. M. K., et al. (2001). "Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance." Nature **414**(6866): 883-887.

Wan, Y., et al. (2011). "Understanding the transcriptome through RNA structure." Nature Reviews Genetics **12**.

Wan, Y., et al. (2013). "Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing." Nature Protocols **8**.

Wan, Y., et al. (2014). "Landscape and variation of RNA secondary structure across the human transcriptome." Nature **505**.

Wang, D. T. and V. J. Dam (2004). "Optical biopsy: A new frontier in endoscopic detection and diagnosis." Clin Gastroenterol Hepatol. **2**(9).

Watrob, H. M., et al. (2003). "Two-Step FRET as a Structural Tool." Journal of the American Chemical Society **125**: 7336-7343.

Weber, P. C., et al. (1989). "Structural Origins of High-affinity Biotin Binding to Streptavidin." Science **243**(4887): 85-88.

Yaroslavsky, A. and S. IV (2013). "Fluorescence Imaging of Single-Copy \ DNA\ Sequences within the Human Genome Using PNA-Directed Padlock Probe Assembly." Chemistry & Biology **20**(3): 445 - 453.

Yildiz, A., et al. (2003). "Myosin V Walks Hand-Over-Hand: Single Fluorophore Imaging with 1.5-nm Localization." Science **300**(5628): 2061-2065.

Zhao, J., et. al. (2010). "Non-B DNA structure-induced genetic instability and evolution." Cell Mol Life Sci **67**(1).

Zhou, Y., et al. (2013). "Recognition of RNA duplexes by chemically modified triplex-forming oligonucleotides." Nucl. Acids Res. **41**(13).

Biography

Vishwa Nellore was born in Hyderabad, India, in 1986. She received her B.E. from Osmania University, India, in 2008, a Master's degree from Rice University, TX, in 2010 and a Ph.D in Electrical and Computer Engineering from Duke University in 2014. She was the recipient of the Pratt Gardner Jr. Fellowship at Duke University and received full scholarship while pursuing her Ph.D. She specializes in the statistical modeling and experimental verification of resonance energy transfer networks constructed using DNA self-assembly. She used the self-assembled networks to construct a novel unclonable key for security as well as to create an RNA structure based sensor that can detect and potentially treat lung and breast cancer. She is passionate about finding a cure for cancer and will continue to work in cancer genomics during her postdoctoral career where she will attempt to identify driver mutations in 12 different cancer types. She has given invited and contributed talks at many conferences including an invited talk titled 'DNA Self-Assembled Cryptographic Device for Two-Factor Authentication' at the Foundations of Nanoscience conference in 2013 and was chosen to give a talk titled 'Using Fluorescence Resonance Energy Transfer Networks to Detect Molecular Signatures in Tumors' at the 12th International Nanomedicine and Drug Delivery Conference in 2014, on winning the best research poster award on the same topic.